Study of VoIP Network Delay Using Neural Networks

Enthrkandi Narasimhan Ganesh

(Corresponding author: E. N. Ganesh)

Department of Electronics & Communication Engineering,
Vels Institute Of Science, Technology & Advanced Studies (VISTAS)
PV Vaithiyalingam Rd, Velan Nagar, Krishnapuram, Pallavaram, Chennai, Tamil Nadu 600117, India
(Email: enganesh50@gmail.com)

(Received Apr. 10, 2019; Revised and Accepted Nov. 20, 2019; First Online June 1, 2020)

Abstract

The statistical nature of data traffic and the dynamic routing techniques employed in IP networks results in a varying network delay (jitter) experienced by the individual IP packets which form a VoIP flow. As a result voice packets generated at successive and periodic intervals at a source will typically be buffered at the receiver prior to playback in order to smooth out the jitter. However, the additional delay introduced by the playout buffer degrades the quality of service. Thus, the ability to forecast the jitter is an integral part of selecting an appropriate buffer size. This paper compares several neural network based models for adaptive playout buffer selection and in particular a novel combined wavelet transform/neural network approach is proposed. The effectiveness of these algorithms is evaluated using recorded VoIP traces by comparing the buffering delay and the packet loss ratios for each technique. In addition, an output speech signal is reconstructed based on the packet loss information for each algorithm and the perceptual quality of the speech is then estimated using the PESQ MOS algorithm. Simulation results indicate that proposed Haar-Wavelets-Packet MLP and Statistical-Model MLP adaptive scheduling schemes offer superior performance.

Keywords: Neural Networks; Playout Delay; Time Series Forecasting And Wavelets; VoIP

1 Introduction

In recent years Voice over IP (VoIP) has seen a huge increase in use due to its cost effectiveness, support of multimedia technology and ease of use. However, the network delay and packet loss, which are ubiquitous due to the best-effort mechanism on which significant portions of the internet are still based are the main factors affecting the Quality of Service (QoS) of a VoIP call [1]. When audio packets are transmitted over the internet, the variable network delay (jitter), which is mainly due to the variable queuing time in routers, modifies the periodic form of the transmitted audio packets when these packets are observed at the receiver [1] as is shown in Figure 1. The playout delay process is an application which aims to reduce the impact of network delay variability by buffering the received packets and playing them out after a certain time. Any packets which arrive later than their playout delay time are regarded as ‘lost packets’ and hence are not played out. Increasing the playout delay can reduce the packet loss, but a long playout delay has a negative impact on the real-time communication quality. Thus, a trade-off exists between the playout delay and packet loss rate. For interactive audio, a packet delay (due to all contributors of delay) of up to 400ms [2] and packet loss rate less than 5% are considered adequate [3]. In early VoIP system, a fixed playout delay was proposed as an initial solution to this problem [4]. While this method offers an easily implemented solution, it is not an optimum solution as it does not take into account the fact that network jitter varies with time, as illustrated in Figure 1.

Modern VoIP systems utilise adaptive playout delay approaches which estimate the network jitter continuously and dynamically adjusts the playout delay at the beginning of each talkspurt. Many algorithms have been proposed for estimating the network jitter such as Autoregressive (AR) models [5], Moving Average (MA) models [6], other statistical models [7–10], and adaptive filter models [11, 12]. In this paper, two new approaches based on combining Artificial Neural Network (ANN) and wavelet techniques and Artificial Neural Network (ANN) and Statistical Models are presented. Zinan lin in [21] and Li etal in [22] uses VOIP for cryptography and steganography.. Hasaneen etal in [23] described in detail about voip in secondary networks for spectrum accessing, Chakraborty explains voip implementation in two tier cognitive network in [24] and equally V.Kumar and trivedi proved the capacity of voip over the same network in [25]. Shintre gave explanation about leakage possibility in [26] and H yang etal described the VOIP over LSTM Network in [27]. Ilyas Khudhair Dubi, Ghiath Mageb Waheeb descrined VOIP Detection over phone and its utility in [28]. Paramjit singh in [29] performed excellent modeling of VOIP network using Fuzzy and kailash Chandra describes in detail on QOS Over VOIP using AI Techniques.

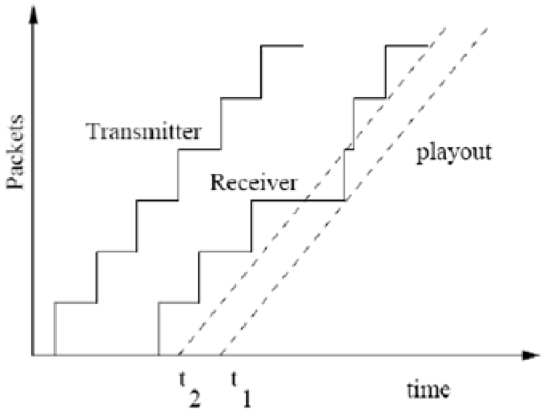


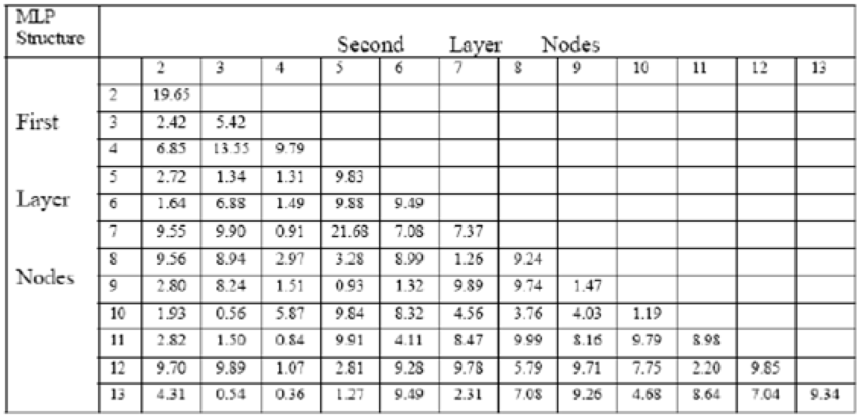
Figure 1: Voice packets over network [4]

2 The Proposed Models

The initial phase of this research focused on the use of neural networks as methodologies for the prediction of network delay jitter. Three types of neural network were compared for forecasting the network delay jitter in the network The first two types are based on a standard multi-layer perceptron (MLP) and a recurrent-MLP [11] with standard training and cross validation methodologies been used to optimize network structure and performance. The last network utilizes a wavelet transform as an input stage prior to applying the resultant signal to an MLP thus forming a Wavelet Packet-MLP (WP-MLP) [12]. Later research being presented focuses on utilizing neural networks to predict the parameters of statistical models of the jitter waveform as an alternative approach.

The traditional back-propagation algorithm using Levenberg-Marquadt with cross validation has been used to train the networks [11]. The data is split into three different data sets used for training, validation (used for cross-validation and structure determination) and testing (used to compare each model after training). Each of the networks has two hidden layers and two outputs. To determine a suitable structure for the network (i.e. the number of nodes in each layer), different network structures were trained (ranging from a 2 × 2 to a 13 × 13 network) and their Prediction Mean Squared Errors (PMSE) compared over the validation set. The best structure was then selected for further evaluation. According to the PMSE performance on the validation set as shown is Table 1 below, the best performing structure was a 10 × 3 network with MSE 5.6 × 10-6.

Table 1: PMSE of Different MLP Structures (× 10 - 5)



MLP Structure		Second Layer Nodes											
		2	3	4	5	6	7	8	9	10	11	12	13
	2	19.65											
First	3	2.42	5.42										
	4	6.85	13.55	9.79									
	5	2.72	1.34	1.31	9.83								
Layer	6	1.64	6.88	1.49	9.88	9.49							
	7	9.55	9.90	0.91	21.68	7.08	7.37						
	8	9.56	8.94	2.97	3.28	8.99	1.26	9.24					
Nodes	9	2.80	8.24	1.51	0.93	1.32	9.89	9.74	1.47				
	10	1.93	0.56	5.87	9.84	8.32	4.56	3.76	4.03	1.19			
	11	2.82	1.50	0.84	9.91	4.11	8.47	9.99	8.16	9.79	8.98		
	12	9.70	9.89	1.07	2.81	9.28	9.78	5.79	9.71	7.75	2.20	9.85	
	13	4.31	0.54	0.36	1.27	9.49	2.31	7.08	9.26	4.68	8.64	7.04	9.34

2.1 Wavelet-Packet Neural Network

In recent years, wavelet networks for function approximation and more specifically time series forecasting has been proposed in [12] and [13]. The wavelet transform maps a time domain signal into a time-frequency domain signal in which the coefficients represent the signal at progressively smaller frequency bands covering larger time spans. Specifically, given a discrete time series $x(k)$ the wavelet transform projects this series onto a new domain known as a wavelet basis [14], as

$$\begin{aligned} x(k) &= \sum_{i=0}^{\infty} \sum_{j=-\infty}^{\infty} W_{i,j}^{v} \Psi_{i,j}(K) && (1) \\ w_{i,j}^{k}(k) &= \int_{-\infty}^{\infty} y(t) \Psi_{i,j}^{*}(k) dt && (2) \\ \Psi_{i,j}(k) &= a_0^{i/2} \Psi(a_0^{i} t - k\tau_0) && (3) \end{aligned}$$

where $t\Psi$ is called the mother wavelet and $\Psi_{i,j}(K)$ is defined in terms of dilations (expansion), a_0 defines a translations (phase shift), τ_0 defines a mother wavelet, and $*$ denotes the complex conjugate. There are various types of mother wavelet, such as Haar wavelet, Meyer Wavelet, Coiflet wavelet, Daubechies wavelet, etc. [14] with the Haar wavelet and Daubechies wavelet being used in this work. After transforming a time series, coefficients which 'contain less information' may be eliminated (shrinkage). This is achieved here by using the variance of the coefficients as a measure of information [14]. When combined with a neural network the overall model is known as a WP-MLP as shown below in Figure 2.

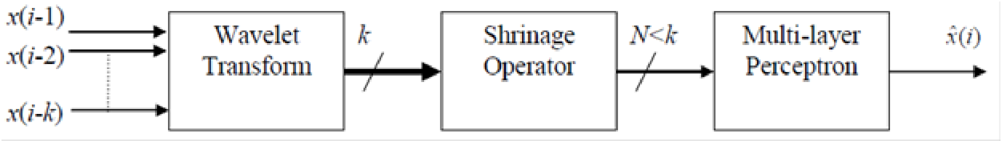


Figure 2: WP-MLP Architecture

2.2 Neural Network Based Statistical Modeling

Several researchers have developed complex models and performed empirical studies of network jitter including those in [15, 16]. These studies show that network jitter follows a Laplacian distribution or a Normal distribution. The probability density function of the normal distribution is

$$
\begin{aligned}
f(x,\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\Pi}} e^{(x-\mu)^2/2\sigma^2} && (4) \\
F(x,\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\Pi}} \int_{-x}^{x} exp(-\frac{(\mu-\mu)^2}{2\sigma^2})du \\
&= \frac{1}{2}91 + erf(\frac{x}{\sqrt{2}})]. && (5)
\end{aligned}
$$

In the proposed technique a neural network is used to predict the mean and variance parameters of a normal distribution model which is then used to calculate the desired playout delay value (ted), as shown in Figure 3.

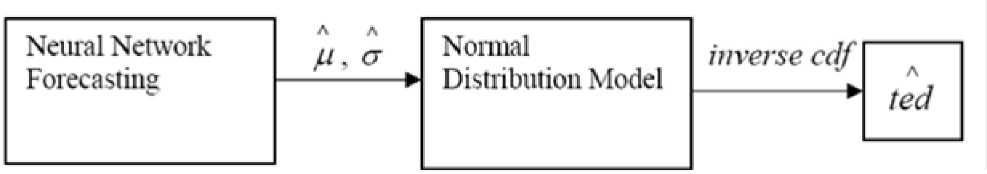

Neural Network Forecasting
$\hat{\mu}, \hat{\sigma}$
Normal Distribution Model
inverse cdf
$\hat{ted}$

Figure 3: Statistical-MLP Modeling

For a chosen mlp (maximum late packet loss percentage desired by user/application) value, there is a corresponding ted delay with any packets experiencing a jitter greater than ted being late for playout and hence discarded. For playout delay adaptation, the ted is chosen as the value of buffering delay that satisfies the condition $1 - cdf(ted) = mlp$.

$$
\begin{aligned}
\frac{1}{2}[1 + erf(\frac{ted}{\sqrt{2}}) &= 1 - m/p && (6) \\
ted &= \sqrt{2}F^{-1}[2(1-m/p)-1] && (7)
\end{aligned}
$$

3 Evaluation Methodology

In this paper, the various models were evaluated using real VoIP traces which were gathered using PJSIP [17], an open source VoIP application written in C, was adapted to measure the network jitter between two hosts. The application used in this paper first encodes the audio stream using G.729 B [18] into 20ms packets of length 80 bytes. Real Time Transport Protocol (RTP) is then used to sequence the packets and these are then encapsulated into a UDP packet for transmission across the internet. Since it was not feasible to take traces using terminals whose clocks were accurately synchronised, only information concerning inter-packet arrival times was available for these traces. Several traces on international VoIP connections where taken ranging in duration from 5 to 10 hours of continuous duplex transmission from NUI, Galway to Tokyo (trace 1, a sample of which shown in Figure 4 below), Sydney (trace 2) and Chengdu (trace 3).

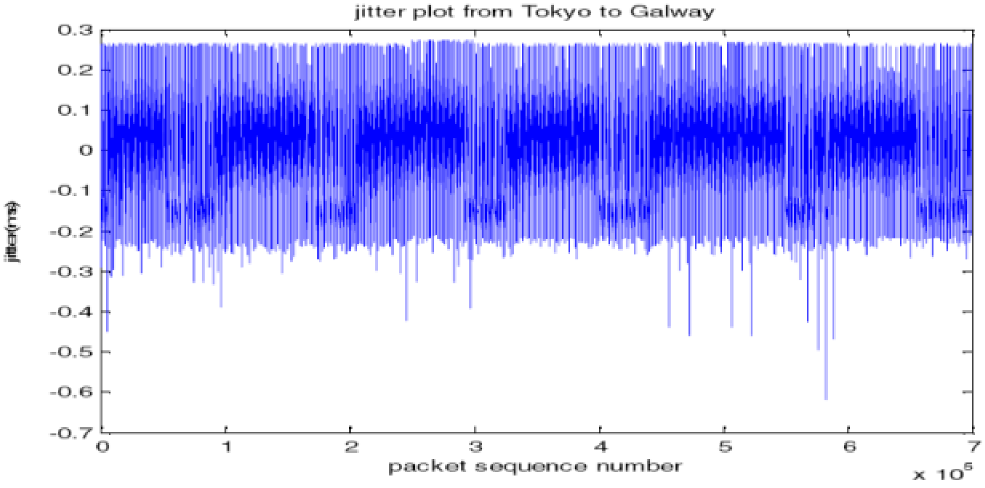


Figure 4: Sample Jitter Plot

At the receiver, an estimate of buffering delay (defined below) is used to allocate the playout time for each talkspurt as:

$$P^k(1) = r^k(1) + ted \tag{8}$$

$$P^k(i) = P^k(1) + (i-1)0.02 i \neq 1, \tag{9}$$

where $P^k(1)$ is the playout time for first packet of the $P^{k^{th}}$ talkspurt, $r^k(1)$ is the arriving time for first packet of the $r^{k^{th}}$ talkspurt, ted is the estimated buffering delay of the $kted\wedge^{th}$ talkspurt, $P^k(i)$ is the playout time for packet i of the $P^{k^{th}}$ talkspurt and 0.02 seconds is ideal interval of the packet playout. $P^k(1)$ is the estimated according to the relative arriving time jitterted Δ_i. The relative arriving time jitter of packet Δ_j^k is defined as:

$$\Delta_j^k(i) = r^k(i) - T^k(i) \tag{10}$$

$$T^k(i) = r^k(1) + (i-1)x_0(i \neq 1) \tag{11}$$

where $\Delta_j^k(i)$ is the relative jitter for packet i of the k^{th} talkspurt, (i) is the arrival time packet i of the $kkri^{th}$ talkspurt, $T^k(i)$ is the ideal arrival time packet i of the $T^k(i)^{th}$ talkspurt with no jitter.

Packets that arrive before their playout time slot () are decoded using G.729 B. Packets that fail to arrive on time or that are dropped are ignored and are decoded instead using the G729 embedded Packet Loss Concealment (PLC) algorithm [19]. This algorithm attempts to interpolate the speech signal using previous packets in the stream. The performance of each proposed model has been analyzed by three metrics:

1) Packet loss rate (the ratio of packets received, prior to, to those sent);

2) PESQ MOS metric; and

3) Additional buffering delay:

$$pd(i) = P^k(i) - r^k(i). \tag{12}$$

Perceptual evaluation of speech quality (PESQ) is a standard to measure the voice quality as published by the ITU-T. It compares a degraded speech signal, which is reconstructed after the network transmission and decoding, to an original signal and a MOS (mean opinion score) value is then produced. Commonly, the MOS value ranges from 0.0 (worst) to 4.5 (best) [20]. The overall algorithm evaluation that was used in the research scheme is shown below in Figure 5.

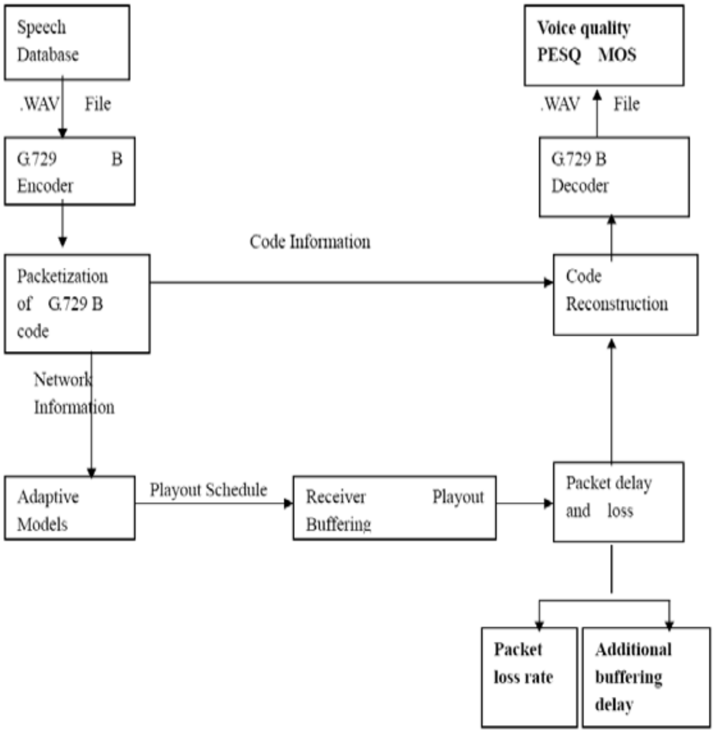


Figure 5: Block Diagram for Performance Analysis Methodology

4 Results

When critically evaluating the performance of a playout scheduling algorithm, it is essential to consider both the additional delay and packet loss rate as shown in Figure 6. This figure illustrates clearly the trade-off between additional delay and packet loss rates for the four different methods being proposed. The P-MLP Haar based algorithm performs best in terms of packet loss (less than the limit for interactive audio, 5% [3]) and additional playout delay up to 400ms [2], which has been improved compared with the raditional MLP. The MLP also shows a good performance, compared with other methods. Comparatively, the results indicate that the RMLP based approach is not very suitable to be used in adaptive playout delay estimation. Alternatively, the Statistical-MLP model also shows a good performance and is very close to the WP-MLP Haar in terms of its abilities.

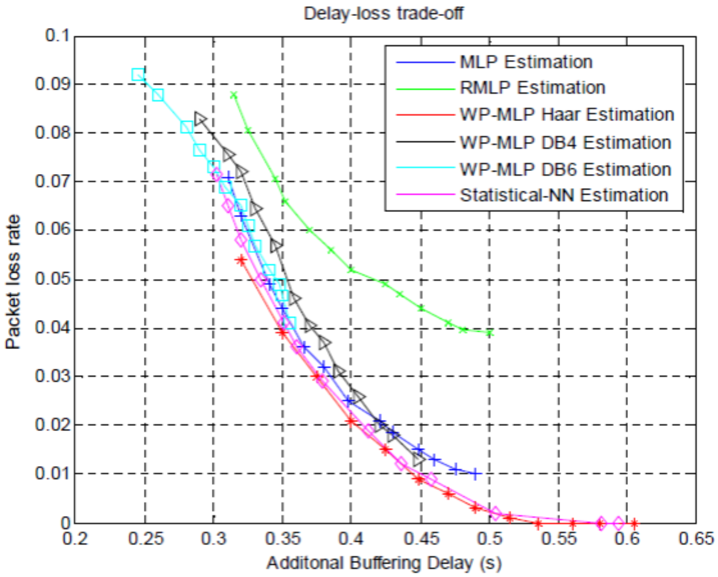


Figure 6: Trade-off between Additional Buffering Delay and Packet Loss Rate

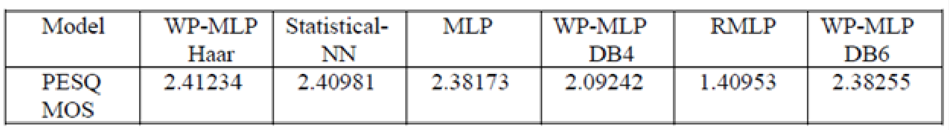
4.1 Results of PESQ MOS based Analysis

Table 2 below, gives a summary of the PESQ MOS score for the five techniques which were evaluated. These PESQ MOS scores were calculated by different lost packet information with the same additional buffering delay of 0.35s. The results show that the WP-MLP using the Haar wavelet also achieved the best performance when considering this perceptual based metric

5 Conclusions

In this paper, several adaptive playout algorithms based on neural network have been presented and their performances evaluated. The effectiveness of these algorithms is evaluated using recorded traces by comparing the buffering delay and the packet loss ratios of each technique. Simulation results indicate that a Haar-Wavelets-Packet MLP adaptive scheduling scheme offers the best performance and

Table 2: Comparison of Algorithm Performance using PESQ MOS Metric



Model	WP-MLP Haar	Statistical-NN	MLP	WP-MLP DB4	RMLP	WP-MLP DB6
PESQ MOS	2.41234	2.40981	2.38173	2.09242	1.40953	2.38255

flexibility for the process of adaptive playout delay estimation. A Statistical Modelling-MLP also shows a good performance very close to that of the WP-MLP (Haar). The WP-MLP DB4 and DB8 models also show an ability to minimise additional buffering delay but at the expense of higher packet loss rates. Future work will focus on the potential for improving the performance of the prediction performance of the WP-MLP by use of different mother wavelets and different levels of decomposition and the statistical-modelling Neural Network which may be improved by considering different combinations of statistical models and neural networks.

References

[1] J. Davidson, J. Peters, *Voice over IP Fundamentals*, Cisco Press, 2000.

[2] Telecommunication Standardization Sector of ITU, *ITU-T Recommendation G.114*, Technical report, International Telecommunication Union, 1993.

[3] N. S. Jayant, "Effects of packet loss on waveform coded speech," in *Fifth Int. Conference on Computer Communications*, Atlanta, Ga., pp. 275–280, 1980.

[4] F. Alvarez-Cuevas, M. Bertran, F. Oller, J. Selga, "Voice Synchronization in Packet Switching Networks," *IEEE Network Magazine*, vol. 7, pp. 20–25, 1993.

[5] R. Ramjee, I. Kurose, D. Towsley, H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," in *IEEE INFOCOM*, pp. 680–688, 1994.

[6] V. Ramos, C. Barakat, E. Altman, "A moving average predictor for playout delay control in VoIP. Quality of Service," *11th International Workshop of IWQoS*, pp. 155–173, 2003.

[7] Y. J. Liang, N. Farber, B. Girod, "Adaptive playout scheduling and loss concealment for voice communications over IP networks," *IEEE Transaction on Multimedia*, vol. 5, pp. 532–543, 2003.

[8] S. B. Moon, I. Kuruse, D. Towslcy, "Packet audio playout delay adjustment: Performance bounds and algorithms," *ACM Multimedia Systems*, vol. 6, pp. 17-28, 1998.

[9] J. Pinto, K. J. Christensen, "An algorithm for playout of packet voice based on adaptive adjustment of talkspurt silence periods," in *IEEE Conf Local Computer Networks*, Lowell, MA, pp. 224-231, 1999.

[10] P. Agrawal, I. C. Chen, C. J. Sreenan, "Use of statistical methods to reduce delays for media playback buffering," in *IEEE Int. Conf Multimedia Computing and Systems*, Austin, TX, pp. 259-263, 1998.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice. Hall, Upper Saddle River, NJ, 1999.

[12] Q. Zhang, A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 889–898, 1992.

[13] M. T. Hagan, M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Transactions on Neural Networks*, vol. 5, pp. 989-993, 1994.

[14] D. B. Percival, A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press, Cambridge, 2000.

[15] L. Zheng, L. Zhang, D. Xu, "Characteristics of Network Delay and Delay Jitter and its Effect on Voice over IP (VoIP)," in *IEEE International Conference on Communications (ICC'01)*, vol. 1, pp. 122-126, 2001.

[16] M. P. Li, J. Wilstrup, R. Jessen, D. Petrich, "A new method for jitter decomposition through its distribution tail fitting," in *ITC Proceeding*, pp. 788-794, 1999.

[17] PJSIP Homepage, May 31, 2020. (`http://www.pjsip.org/`)

[18] ITU, *ITU-T Recommendation G.729 Annex B., A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70*, 1996.

[19] ANSI, *Packet Loss Concealment for use with ITU-T Recommendation G.711. ANSI Recommendation T1.521a-2000 (Annex B)*, 2000.

[20] ITU, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, International Telecommunication Union, 2001.

[21] Z. Lin, Y. Yuag, "Fast steganalysis of VOIP streams using Recurrent Neural Networks," *IEEE Transaction on Information Forensics and Security*, vol. 13, pp. 1854–1868, 2018.

[22] S. Li, Y. Jia, C. C. J. Kuo, "Steganalysis of QIM steganography in low-bit-rate speech signals," *IEEE/ACM Trans. Audio Speech Language Process*, vol. 25, no. 5, pp. 1011-1022, 2017.

[23] H. S. Hasaneen, "Secondary VOIP Capacity in Opportunistic Spectrum access network with Friendly scheduling," *IEEE Transaction on Mobile Computing*, vol. 15, no. 3, pp. 733-747, 2016.

[24] T. Chakraborty, "Design and Implementation of VoIP Based Two-Tier Cognitive Radio Network for Improved Spectrum Utilization," *IEEE Systems Journal*, 2016. (DOI:10.1109/jsyst.2014.2382607)

[25] V. Kumar, A. Trivedi, "Capacity improvement for VoIP based two-tier CRN using space-time spectrum sensing," in *Proceedings of International Conference on Advances in Computing*, 2016. (DOI:10.1109/icacce.2016.8073740)

[26] S. Shintre, V. Gligore, Ja Ao Borros, "Anomity leakage in private VOIP Networks," *IEEE Transaction on Dependable and Secure Computing*, vol. 15, no. 1, pp. 14-26, 2018.

[27] H. Yang, Z. Yan, Y. Young, "Steganalysis of VoIP Streams with CNN-LSTM Network," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, vol. 1, pp. 204-209, 2019.

[28] I. K. Dubi, G. M. Waheeb, "System for Rogue VoIP Phone Detection and Managing VoIP Phone Mobility," *International Journal of Recent Trends in Engineering*, vol. 8, no. 2, pp. 2469-2474, 2019.

[29] P. Singh, A. K. Sharma, T. S. Kamal, "An adaptive neuro-fuzzy inference system modeling for VoIP based IEEE 802.11g MANET," *Optik*, vol. 127, no. 1, pp. 122-126, 2016.

[30] K. Chandra and P. C. Das, "Measuring Quality of Service of VoIP Based on Artificial Neural Network Approach," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 3, pp. 657-661, 2015.

Biography

Dr. E. N. Ganesh Professor in the field of Electronics for the past 20 years, Specialised in Nanoelectronics and Microelectronics. His research work in Quantum electronics adjudged best thesis and received Gold medal in Phd. He has finished M.Tech from IIT Madras in Microelectronics. He has 54 Conferences and Journal Publications.