International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)
ISSN: 2305-0543 (Online), ISSN: 2411-6173 (Print)
Vol. 11(39), Jan. 2021, PP. 4860-4868

Comparative Performance Evaluation of Voice Coding Schemes in Bandwidth-Constrained VoIP Networks

Justus N. Dike
Dept. of Electrical/Electronic Engineering
University of Port Harcourt
Port Harcourt, Nigeria
justus.dike@uniport.edu.ng

Cosmas I. Ani
Dept. of Electronic Engineering
University of Nigeria
Nsukka, Nigeria
cosmas.ani@unn.edu.ng

Abstract — In **Voice over Internet Protocol (VoIP) networks, bandwidth consumption depends on the coding scheme or encoding algorithm used. The lower the coder/decoder (CODEC) bit rate, the less bandwidth is required for transmission, leading to a more efficient system. On the other hand, the higher the bit rate of the encoded bit stream, the higher the voice quality but at a higher cost. That is, more bandwidth or transmission capacity is required. However, the greater the savings in transmission data rate (that is, lower bit rate), the worse the perceived quality as a result of the longer time it takes to encode the voice samples. Hence, a trade-off is necessary to satisfy the demands of a particular application. This paper presents the performance evaluation comparison of the G.711, G.726 and G.729 coding schemes in an optimized quality of service (QoS)-based packet scheduling algorithm for bandwidth-constrained VoIP networks. The design of an optimized packet scheduling model as well as an analytical appraisal of the developed model were carried out. Riverbed (formally OPNET) Modeler version 17.5 was used to simulate the developed model. Simulation results for CODEC end-to-end delay performance gave 34.19%, 33.86% and 31.94% respectively for G.711, G.726 and G.729. Packet loss probability performance gave 60%, 33.33% and 6.67% respectively. Resource utilization performance gave 44.44%, 37.04% and 18.52% respectively. The results obtained show that the G.729 coding scheme, which has the lowest bit rate offers the best performance and therefore guarantees better network QoS performance. This fact validates the optimal performance of the proposed QoS-based model in bandwidth-constrained VoIP networks.**

Keywords — coder, decoder, performance, evaluation, delay, packet-loss, resource, utilization

I. INTRODUCTION

Two critical quality of service (QoS) metrics in Voice over Internet Protocol (VoIP) networks are optimal bandwidth consumption and end-to-end, absolute transmission (mouth-to-ear (M2E)) delay. Owing to the increasing demand for voice communication over the Internet, voice compression technology has become a critical component of optimizing QoS. This is because voice compression conserves system capacity. Voice compression is a procedure to represent a digitized speech using as few bits as possible but maintaining at the same time a reasonable level of voice quality. Every application on the network utilizes a portion of the available bandwidth. Hence, maximizing bandwidth consumption is one of the most critical aspects of network management and one of the major tasks faced by network designers and administrators [1, 2].

In VoIP networks, bandwidth consumption naturally depends on the coding scheme or encoding algorithm (commonly referred to as coder/decoder (CODEC)) used [3]. The lower the CODEC bit-rate, the less bandwidth is required for transmission, leading to a more efficient system. This requirement is in constant conflict with other good properties of the system, such as speech quality [1]. The higher the bit rate of the encoded bit-stream, the higher the voice quality but at a higher cost - more bandwidth or transmission capacity is required. On the other hand, the greater the savings in transmitted data rate (that is, lower bit rate), the worse the perceived quality as a result of the longer time it takes to encode the voice samples [1]. This is because increase in the coder or processing delay increases the overall mouth-to-ear (transmission) delay. A trade-off is therefore necessary to satisfy the demands of a particular application.

The Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) Series G defines the coding of voice and audio signals of digital terminal equipment of transmission systems and media, digital systems and networks. Some of the common schemes employed in VoIP networks are the traditional sample-based G.711 CODEC with a data rate of 64 kbps [1, 4, 5], the G.726 with 16-40 kbps [5, 6], the G.728 with 16 kbps [1, 7, 8]. The more recent frame-based encoders provide drastic rate reduction (for example, 8 kbps for G.729 as well as 5.3 and 6.3 kbps for G.723.1) at the expense of additional complexity and encoding delay as well as lower quality [9, 10]. This paper therefore investigates the impact of voice compression on the quality of service of bandwidth-constrained VoIP networks and presents a comparison of the performance evaluation of G.711, G.726 and G.729 coding schemes employed in an optimized QoS-based packet scheduling algorithm.

II. RELATED WORKS

Several optimization solutions, including [11, 12, 13, 14, 15, 16, 17, 18, 19, 20], employing different approaches have

Article History:
Received Date: Aug. 03, 2020, Accepted Date: Sep. 14, 2020, Available Online: Jan. 01, 2021

addressed the QoS issues of VoIP networks on both wired and wireless topology implementations. Similarly, several proposals, including [21, 22, 23, 24, 25, 26, 27, 28, 29, 30], have investigated the impact of different voice coding schemes on QoS optimization in VoIP networks. Earlier works lumped critical and best-effort data together. In this work, an optimized voice and critical-data QoS-based hybrid packet scheduler is proposed. The paper therefore presents a comparative performance evaluation of common voice coding schemes in bandwidth-constrained VoIP networks.

III. Design of an Optimized QoS-Based Packet Scheduling Model

The architecture of the proposed model, illustrated in Figure 1, is an integration of packet classifiers, Token Bucket, Differentiated Services (DiffServ) and Weighted Round Robin (WRR) modules. The Packet Classifier module comprises two packet classifiers. Classifier1 classifies the packets of the incoming source traffic (p) into two main classes, namely: voice ($p1$) and non-voice ($p2$) flows. Packet Classifier2 classifies the non-voice flows into two other classes, namely: business/mission-critical data (B/MCD, $p3$) and others ($p4$ - consisting of video and remaining (best effort) data traffics). The essence of Classifier2 is to capture and accord B/MCD flows the necessary priority and fairness they deserve.

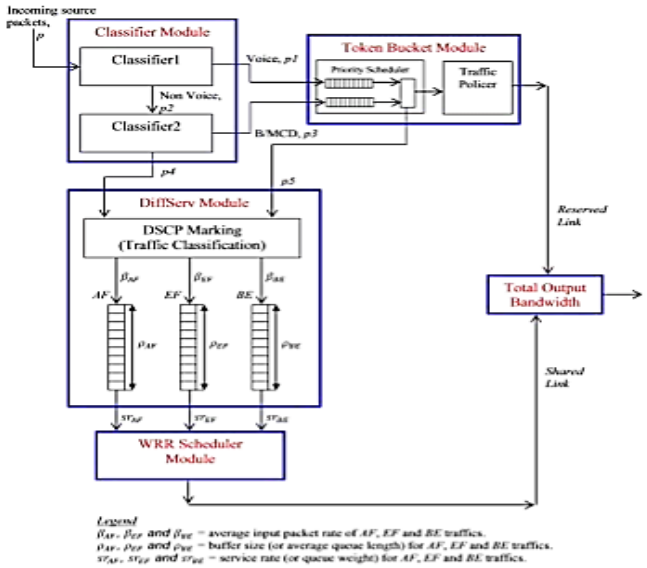

Figure 1: Architecture of the Optimized QoS-Based Packet Scheduler

The Token Bucket module is used to split the incoming voice or B/MCD traffic into two sub-flows. The first sub-flow is a well-shaped flow with maximum rate equal to γ bits/second generated by the Token Bucket. The second sub-flow is the packet ($p5$ - still of voice or B/MCD traffic) rejected by the Token Bucket. Non-preemptive priority scheduling discipline is employed for forwarding voice and B/MCD traffics to the Token Bucket. This implies that there is no interruption to any traffic being transmitted through the Bucket. Voice traffic is classified into the high priority class while B/MCD traffic is classified into the low priority class at the output queue.

In the DiffServ module, video traffic is mapped to Assured Forwarding (AF) traffic class. Voice or B/MCD traffic, which was rejected from the Token Bucket is mapped to the Expedited Forwarding (EF) traffic class. The remaining data traffic (such as email, file transfer, and so on) is mapped by default to the Best Effort (BE) class.

The WRR scheduler module is used to adaptively regulate the bandwidth utilization among the competitive traffic flows from the DiffServ module. The output (constrained) bandwidth is divided into two parts, namely: the reserved (dedicated) link and the shared link. The reserved link is used to service the specified portion of voice or business/mission-critical data traffic from the Token Bucket. The shared link is used to service the other traffics as scheduled fairly and adaptively by the WRR scheduler [31, 32, 33, 34, 35, and 36].

IV. Analytical Appraisal of the Developed Model

The approximation method of analysing queuing networks by decomposition is hereby adopted. It is the decomposition of the whole network or the aggregation of portions of the network. The arrival rate is assumed to be Poisson and the service times of network elements are exponentially distributed in this method. The steps involved in the analysis by decomposition is given as follows [32]:

(a) Isolate the queueing network into subsystems (such as single servers or transmission links).
(b) Analyze each subsystem separately, considering its own network surroundings of arrivals and departures.
(c) Find the average delay and packet-loss probability for each individual queueing subsystem, and
(d) Aggregate all the delays and packet-loss probabilities of queueing subsystems to find the average total end-to-end network delay and packet-loss probability.

The offered or traffic load (traffic intensity or server utilization factor, which is the load rate at which packets arrive) at the network is defined as:

$$\rho_{net} = \frac{\beta_{net}}{\mu_{net}} = \frac{Average\ service\ completion\ time\ (1/\mu_{net})}{Average\ interarrival\ time\ (1/\beta_{net})} \quad (1)$$

A critical optimal network QoS requirement demands that the offered load (ρ_{net}) should be less than 1. This implies that the arrival rate (β_{net}) should not be allowed to exceed the maximum capacity (μ_{net}) of the network. This work therefore ensures adequate bounding in the in-built congestion control mechanisms of the proposed model, which ensures that $\beta_{net} < \mu_{net}$ ($\rho_{net} < 1$).

The total number of packets in a given network (N_{net}), by employing Little's Theorem [19, 20], is defined as:

$$Av[N_{net}] = \beta_{net}(1 - P_{pl-net})Av[D_{net}] \quad (2)$$

where β_{net} is the total packet arrival rate to the network, P_{pl-net} is packet-loss probability and D_{net} is the

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

sum of the waiting times and servicing times experienced by the packet in the network. $\beta_{net}(1 - P_{pl-net})$ is the actual packet arrival rate into the network.

In a network comprising several transmission links in several domains, equation (2) can be rewritten for every transmission link in every domain. Hence, for the k^{th} transmission link, the average number of packets is given by:

$$Av[N_k] = \beta_k(1 - P_{pl-k})Av[D_k] \quad (3)$$

where β_k is the packet arrival rate at the k^{th} transmission link, P_{pl-k} is the packet-loss probability and $Av[D_k]$ is the average delay experienced by the packet in the link.

The average of the total number of packet in the network is therefore equal to the sum of the average number of packets in all the transmission links in the network and is given by:

$$Av[N_{net}] = \sum_k Av[N_k] = \sum_k \left[\beta_k(1 - P_{pl-k})Av[D_k]\right] \quad (4)$$

Hence, the average total delay ($Av[D_{net}]$) experienced by packets in transversing the entire network is derived from equation (2) as:

$$Av[D_{net}] = \frac{1}{\beta_{net}(1 - P_{pl-net})} Av[N_{net}] = \frac{1}{\beta_{net}(1 - P_{pl-net})} \sum_k \left[\beta_k(1 - P_{pl-k})Av[D_k]\right] \quad (5)$$

Equation (5) therefore states that the total network delay depends on the actual packet arrival rate to the network ($\beta_{net}(1 - P_{pl-net})$) as well as the actual packet arrival rate ($\beta_k(1 - P_{pl-k})$) to, and the delay (D_k) in every transmission link in every domain constituting the network [32].

Figure 2 illustrates the state transition diagram for determining the probability (P_n) of n number of packets in the network at time t. The number increases by 1 in the next small interval of time Δt seconds (just before time, t) with probability $\beta\Delta t$ and decreases by 1 in the next Δt seconds with probability $\mu\Delta t$ [34, 39, 40].

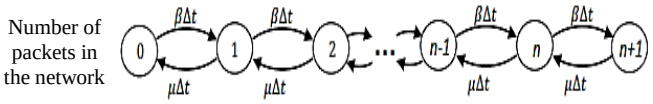

Figure 2: State transition diagram [21, 22]

In the single-server queuing model, network stability is attained when the long-run transition rate ($P_n\beta\Delta t$) from state n to state $n + 1$ equals the long-run transition rate ($P_{n+1}\mu\Delta t$) from state $n + 1$ to state n. This implies that [34, 39, 40]:

$$P_{n+1} = \left[\frac{\beta}{\mu}\right] P_n, \quad for\ n = 0, 1, \dots, K = \left[\frac{\beta}{\mu}\right]^{n+1} P_0 \quad (6)$$

The probability of being in state n (that is, the probability that n packets are in the network), which is also the proportion of time that the network is in state n is defined by:

$$P_n = \left[\frac{\beta}{\mu}\right]^n P_0, \quad for\ n = 0, 1, \dots, K = \rho^n P_0 \quad (7)$$

where P_0 is the proportion of time that the network is empty and K is the maximum occupancy (or capacity) of the network.

Applying the normalization condition where all probabilities add up to 1 gives:

$$1 = \sum_{n=0}^{K} P_n = P_0(1 + \rho + \rho^2 + \rho^3 + \dots \rho^K) = P_0\left[\frac{1-\rho^{K+1}}{1-\rho}\right] \quad (8)$$

$$\gg\gg \quad P_0 = \left[\frac{1-\rho}{1-\rho^{K+1}}\right] \quad (9)$$

Hence, the probability for the number of packets in the network is given by:

$$P[N_{net}(t) = n] = P_n = {\rho_{net}}^n \left[\frac{(1-\rho_{net})}{(1-{\rho_{net}}^{K+1})}\right], for\ n = 0, 1, \dots, K \quad (10)$$

The average number of packets in the network is given by:

$$\boldsymbol{Av[N_{net}] = \sum_{n=0}^{K} nP_n} = \sum_{n=0}^{K} \boldsymbol{n}\,{\rho_{net}}^{\boldsymbol{n}} \left[\frac{(\boldsymbol{1-\rho_{net}})}{(\boldsymbol{1-{\rho_{net}}^{K+1}})}\right] = \left\{\left[\frac{\rho_{net}}{(1-\rho_{net})}\right] - \left[\frac{(K+1){\rho_{net}}^{K+1}}{(1-{\rho_{net}}^{K+1})}\right]\right\} \quad (11)$$

The packet-loss probability of the network, which is also the proportion of time that the network is full, is given from equation (10) by:

$$P_{pl-net} = P_K = {\rho_{net}}^K \left[\frac{(1-\rho_{net})}{(1-{\rho_{net}}^{K+1})}\right] \quad (12)$$

Hence, the probability of the average delay in the network is given by:

$$Av[D_{net}] = \frac{Av[N_{net}]}{\beta_{net}(1-P_{pl-net})} = \left[\frac{\left\{\left[\frac{\rho_{net}}{(1-\rho_{net})}\right] - \left[\frac{(K+1){\rho_{net}}^{K+1}}{(1-{\rho_{net}}^{K+1})}\right]\right\}}{\left\{\rho_{net}\mu_{net}\left[1 - {\rho_{net}}^K\left(\frac{(1-\rho_{net})}{(1-{\rho_{net}}^{K+1})}\right)\right]\right\}}\right] \quad (13)$$

In this analysis, packets are transmitted in one domain (or Service Provider) via two links. The dedicated transmission link services the reserved flow of voice or B/MCD that has an upper bound rate that is equal to $\gamma\ bits/second$ in the Token Bucket module while the shared transmission link services the flow regulated by the WRR scheduler from the DiffServ module. This implies that the average delay and packet loss encountered by traffic flows in each domain are accounted for by these links.

In the Token Bucket module [32, 34], let TP be the total number of packets (played out from the priority scheduler) of voice (TP_V) or B/MCD ($TP_{B/MCD}$). Each of these traffics is divided into two parts, namely: the reserved (or dedicated) and rejected (or surplus) flow. The reserved flow, TP_{rev} (TP_{V-rev} or $TP_{B/MCD-rev}$) has an upper bound rate that is equal to $\boldsymbol{\gamma}\ bits/second$ and is expressed as (αTP), where α is the splitting ratio of the Token Bucket and is defined as

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

$(0 \leq \alpha \leq 1)$. TP_{V-rev} or $TP_{B/MCD-rev}$ is transmitted via the reserved link only if TP_{V-rev} or $TP_{B/MCD-rev}$ is less than (<) or equal to (=) $\boldsymbol{\gamma}\ bits/second$. The rejected flow, TP_{rej} (TP_{V-rej} or $TP_{B/MCD-rej}$) is directed to the DiffServ (DS) module and is expressed as$[(1-\alpha)TP]$. A splitting ratio of $\alpha = 0$ implies that all the packets are scheduled via the WRR scheduler only (that is, $TP_{rev} = 0$). A splitting ratio of $\alpha = 1$ implies that $TP_{rej} = 0$. The splitting ratio of the token bucket is used to ensure that the reserved link is never redundant at any time, t. That is, even if TP_V or $TP_{B/MCD}$ is greater than (>) $\boldsymbol{\gamma}\ bits/second$, the splitting ratio is used to ensure that TP_{V-rev} or $TP_{B/MCD-rev}$ is less than (<) or equal to (=) $\boldsymbol{\gamma}\ bits/second$. These imply that:

$$TP_{rev}\left(TP_{V-rev}\ or\ TP_{B/MCD-rev}\right) = \gamma = (\alpha TP)$$
$$TP_{rej}\left(TP_{V-rej}\ or\ TP_{B/MCD-rej}\right) = [(1-\alpha)TP] \quad (14)$$

Let us now consider a typical run of the proposed hybrid scheduling model during a time interval $(0, t)$ of the run. Since the voice and B/MCD traffic flows are each divided into two sub flows, the rate of the first sub flow (αTP) will never exceed the token bucket rate γ bits/second. Hence, the upper bound of bits serviced by this flow in the time interval $(0, t)$ is equal to $\boldsymbol{\gamma}(0, t)$. By effectively varying the splitting ratio, adequate precedence to both voice and B/MCD traffic workloads as well as bandwidth utilization fairness are ensured in the network. Applying Little's Theorem (equation 2) in the Token Bucket gives:

$$Av[N_{TB}] = \beta_{TB}(1 - P_{pl-TB})Av[D_{TB}]$$
$$\gg\gg \qquad Av[D_{TB}] = \frac{1}{\beta_{TB}(1-P_{pl-TB})}Av[N_{TB}] \quad (15)$$

where $Av[N_{TB}]$ is the average number of packets in the bucket, β_{TB} is the arrival rate (that is, the average number of packets arriving the bucket per unit time) and $Av[D_{TB}]$ is the average delay (or time spent) in the bucket. $\beta_{TB}(1 - P_{pl-TB})$ is the actual packet arrival rate into the bucket.

This implies that for a run time $(0, t)$, the average time spent by the average number of packet in the bucket equals the time spent in transmitting the reserved flow through the dedicated link. Hence, for the dedicated link, the state probabilities are defined according to Equations (equations 12 and 13). The average number of packets in the bucket is given by:

$$Av[N_{TB}] = \left\{\left[\frac{\rho_{TB}}{(1-\rho_{TB})}\right] - \left[\frac{(K_{TB}+1)\rho_{TB}{}^{K_{TB}+1}}{(1-\rho_{TB}{}^{K_{TB}+1})}\right]\right\} \quad (16)$$

where $\rho_{TB} = \beta_{TB}/\mu_{TB}$ is the offered load (or load rate at which packets arrive at the bucket); μ_{TB} is the maximum departure rate at which packets are serviced or transmitted via the reserved (or dedicated) link and K_{TB} is the maximum occupancy in the dedicated link.

The packet-loss probability of the reserved link (RL) is defined as:

$$P_{pl-RL} = P_{pl-TB} = \rho_{TB}^{K_{TB}}\left[\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right] \quad (17)$$

Hence, the probability of the average delay in the reserved (or dedicated) link is given by:

$$Av[D_{RL}] = Av[D_{TB}] = \frac{Av[N_{TB}]}{\beta_{TB}(1-P_{pl-TB})}$$

$$= \left[\frac{\left\{\left[\frac{\rho_{TB}}{(1-\rho_{TB})}\right] - \left[\frac{(K_{TB}+1)\rho_{TB}{}^{K_{TB}+1}}{(1-\rho_{TB}{}^{K_{TB}+1})}\right]\right\}}{\left\{\rho_{TB}\mu_{TB}\left[1-\rho_{TB}^{K_{TB}}\left(\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right)\right]\right\}}\right] \quad (18)$$

In the DiffServ module [32, 34], the surplus flow $[(1-\alpha)TP]$ rejected from the token bucket is marked and classified by Expedited Forwarding (EF) Differentiated Service Code Point (DSCP) value [41]. Similarly, video traffic is marked and classified by Assured Forwarding (AF) DSCP value [42] while the remaining data traffic is marked and classified as default by Best Effort (BE) DSCP value. The average input packet arrival rates for the EF, AF and BE traffics are respectively defined as β_{EF}, β_{AF} and β_{BE}. The total available buffer and shared link bandwidth (SLB) are distributed to the different classes of services. The buffer allocation procedure is adaptive with the current offered load, ρ (or the current queue length, ql) to that queue. The SLB procedure uses the buffer size to compute the queue weight (or service rate) in the Weighted Round Robin (WRR) scheduler module. The service rate of each queue is proportional to the allocated bandwidth to that queue. From this background, the allocated buffers (the current offered loads or queue lengths) for the queues of EF, AF and BE classes of traffic are respectively computed as follows [32, 34, 43]:

$$\rho_{EF} = \frac{\beta_{EF} \times \delta_{EF} \times N_{EF}}{PS_{EF}}$$
$$\rho_{AF} = \frac{\beta_{AF} \times \delta_{AF} \times N_{AF}}{PS_{AF}}$$
$$\rho_{BE} = \frac{\beta_{BE} \times \delta_{BE} \times N_{BE}}{PS_{BE}} \quad (19)$$

where: δ_{EF}, δ_{AF} and δ_{BE} are respectively the allowable packet delays; N_{EF}, N_{AF} and N_{BE} are respectively the number of sessions; PS_{EF}, PS_{AF} and PS_{BE} are respectively the average packet sizes for EF, AF and BE traffics.

In the WRR scheduler module [32, 34], the traffic queues are serviced according to the service rate (sr) (or queue weight) of each class of traffic. Service rate is adaptive with the current offered load (ρ) of a particular class of traffic. In any time interval $(0, t)$ of any round robin execution of the proposed model, the policing procedure is such that the traffic with the maximum computed service rate is serviced first. If the computed service rates for the three classes at this particular time interval are equal, the scheduling algorithm uses the assigned priority values. In this proposal, EF traffic has the highest priority value (PV) of 3, followed by AF traffic (2) and then BE traffic (1). The computed service rates (sr) or queue weights [32, 34, 43] for the three classes of traffic are given as follows:

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

$$sr_{EF}(t) = \frac{\rho_{EF}(t) \times PV_{EF}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)}$$
$$sr_{AF}(t) = \frac{\rho_{AF}(t) \times PV_{AF}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)}$$
$$sr_{BE}(t) = \frac{\rho_{BE}(t) \times PV_{BE}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)} \quad (20)$$

where: PV_{EF}, PV_{AF} and PV_{BE} are respectively the priority values for EF, AF and BE traffics at the interval $(0, t)$.

Equation (20) clearly shows that the service rate of a particular service queue is adaptive to the ratio of the buffer size (or length of that queue) to the sum of the buffer sizes (queue lengths) of all service queues at the interval of time $(0, t)$. This implies that the service queue that has the longest queue length is services first, followed by that with the next longest queue length, and so on. By the round robin operation, every service queue present is considered in that order in every round robin execution. As stated earlier, if the service rates of all the three traffic classes are equal, the proposed model employs the priority values (3 for EF, 2 for AF and 1 for BE traffics) in forwarding the service queue to the shared link. Again, if there is no packet present in a particular traffic class at the time interval under consideration, the algorithm forwards the available service queue, and so on.

If the transmission rate of the shared link (SL) is R_{T-SL}, the throughputs of the three classes of traffic are respectively given as follows:

$$Throughput_{EF} = \frac{(R_{T-SL}) \times (sr_{EF})}{sr_{WRR}}$$
$$Throughput_{AF} = \frac{(R_{T-SL}) \times (sr_{AF})}{sr_{WRR}}$$
$$Throughput_{BE} = \frac{(R_{T-SL}) \times (sr_{BE})}{sr_{WRR}} \quad (21)$$

where sr_{WRR} is the service rate of the weighted round robin scheduler and defined by:

$$sr_{WRR} = sr_{EF} + sr_{AF} + sr_{BE} \quad (22)$$

The arrivals at the EF, AF and BE priority classes are Poisson with rates: β_{EF}, β_{AF} and β_{BE} respectively. The average servicing (or transmission) times for the queues are respectively defined as: $Av[S_{EF}] = 1/\mu_{EF}$, $Av[S_{AF}] = 1/\mu_{AF}$ and $Av[S_{BE}] = 1/\mu_{BE}$; where μ_{EF}, μ_{AF} and μ_{BE} are respectively the maximum departure rates for EF, AF and BE traffics. Hence, the respective loads offered by EF (the highest priority class), AF (the next highest priority class) and BE (the least priority class) are given as:

$$\rho_{EF} = \beta_{EF} Av[S_{EF}]$$
$$\rho_{AF} = \beta_{AF} Av[S_{AF}]$$
$$\rho_{BE} = \beta_{BE} Av[S_{BE}] \quad (23)$$

In a typical priority class-based system, if the total offered load for C (number of) priority classes is less than 1, that is:

$$\rho = \rho_1 + \rho_2 + \cdots + \rho_C < 1 \quad (24)$$

then the average waiting time for a type C packet is given by [39]:

$$Av[W_C] = \frac{\beta Av[S^2]}{(1 - \rho_1 - \rho_2 - \ldots - \rho_{C-1})(1 - \rho_1 - \rho_2 - \ldots - \rho_C)} \quad (25)$$

where ρ_1 is the offered load of the highest priority class, and so on.

In this work, $C = 3$. This implies that the total offered load ρ_{WRR} (of the weighted round robin scheduler) is:

$$\rho_{WRR} = \rho_{EF} + \rho_{AF} + \rho_{BE} < 1 \quad (26)$$

and the average waiting times for the EF, AF and BE flows are respectively defined as:

$$Av[W_{EF}] = \frac{\beta_{WRR} Av[S^2_{WRR}]}{(1 - \rho_{EF})}$$
$$Av[W_{AF}] = \frac{\beta_{WRR} Av[S^2_{WRR}]}{(1 - \rho_{EF})(1 - \rho_{EF} - \rho_{AF})}$$
$$Av[W_{BE}] = \frac{\beta_{WRR} Av[S^2_{WRR}]}{(1 - \rho_{EF} - \rho_{AF})(1 - \rho_{EF} - \rho_{AF} - \rho_{BE})} \quad (27)$$

where:

$$Av[S^2_{WRR}] = \frac{\beta_{EF}}{\beta_{WRR}} Av[S^2_{EF}] + \frac{\beta_{AF}}{\beta_{WRR}} Av[S^2_{AF}] + \frac{\beta_{BE}}{\beta_{WRR}} Av[S^2_{BE}] \quad (28)$$

and $$\beta_{WRR} = \beta_{EF} + \beta_{AF} + \beta_{BE} \quad (29)$$

The average delay for each class of flow is determined by adding the average of the service (or transmission) time to the corresponding average waiting time. That is:

$$Av[D_{EF}] = Av[W_{EF}] + Av[S_{EF}]$$
$$Av[D_{AF}] = Av[W_{AF}] + Av[S_{AF}]$$
$$Av[D_{BE}] = Av[W_{BE}] + Av[S_{BE}] \quad (30)$$

The average delay of the shared link, $Av[D_{SL}]$, which is the average delay of the weighted round robin scheduler, is therefore given as:

$$Av[D_{SL}] = Av[D_{WRR}] = \frac{1}{3}\{Av[D_{EF}] + Av[D_{AF}] + Av[D_{BE}]\} \quad (31)$$

The packet-loss probability of the shared link is defined as:

$$P_{pl-SL} = P_{pl-WRR} = \rho_{WRR}^{K_{WRR}} \left[\frac{(1 - \rho_{WRR})}{\left(1 - \rho_{WRR}^{(K_{WRR}+1)}\right)}\right] \quad (32)$$

where ρ_{WRR} is the offered load and K_{WRR} is the maximum occupancy in the shared link.

By aggregating the mean delays of the reserved and shared links, the total average delay of the proposed optimized model is obtained from equation (5) as:

$$Av[D_{MODEL}] = \left[\frac{1}{\beta_{MOD}(1 - P_{pl-MOD})}\{\beta_{TB}(- P_{pl-TB})Av[D_{TB}] + \beta_{WRR}(1 - P_{pl-WRR})Av[D_{WRR}]\}\right] \quad (33)$$

where $\beta_{MOD}(1 - P_{pl-MOD})$ is the total actual packet arrival rate to the optimized model; $\beta_{TB}(1 - P_{pl-TB})$ and D_{TB} are the actual packet arrival rate and delay for the reserved link while $\beta_{WRR}(1 -$

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

P_{pl-WRR}) and D_{WRR} are the actual packet arrival rate and delay for the shared link.

Similarly, by aggregating the packet-loss probabilities of the reserved and shared links, the total packet-loss probability of the proposed model is defined as:

$$P_{pl-MODEL} = P_{pl-RL} + P_{pl-SL} = \left[\left[\rho_{TB}^{K_{TB}}\left(\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right)\right] + \left[\rho_{WRR}^{K_{WRR}}\left(\frac{(1-\rho_{WRR})}{\left(1-\rho_{WRR}^{(K_{WRR}+1)}\right)}\right)\right]\right] \quad (27)$$

(34)

The fundamental QoS requirement that $\rho < 1$ ($\beta < \mu$) is therefore ensured in the reserved link by the traffic regulation of the Token Bucket. Here, the upper bound of γ bits/second must not be allowed to exceed the bandwidth capacity of the reserved link. In the shared link, the traffic regulation of the WRR scheduler also ensures that what plays out at every instant in time within the run time does not exceed the bandwidth capacity. In fact, the computed throughputs (equation 21) show that only a fraction of the capacity of the link is allocated to each aggregated flow.

V. Simulation Model

The proposed QoS-based model is hereby simulated with Riverbed (formally OPNET) Modeler version 17.5 [44, 45]. Figure 3 illustrates the network topology used for the simulation. Nodes n1, n2, n3 and n4 are respectively the voice, business/mission-critical data (B/MCD), video and best-effort data sources (generators). These nodes are connected to the edge switch (n5), which in turn is connected to the edge router (n6), all at the sending end. The edge router is then connected to the network via a constrained (bottlenecked) bandwidth. The connection is similar but reversed at the receiving end. At this end, the edge router connects the edge switch, which in turn connects the voice, B/MCD, video and best-effort data sinks. The link capacity between the data sources/sinks and edge switches is 100Mbps while that between the edge switches and routers is 1Gbps. The capacity of the bottlenecked-link is 2Mbps. The nodes, switches, routers and links were appropriately configured [33, 34, 35]. The end-to-end delay, packet loss, network throughput and resource utilization QoS impairment parameters were monitored at the receiving end using each of the Coding Schemes.

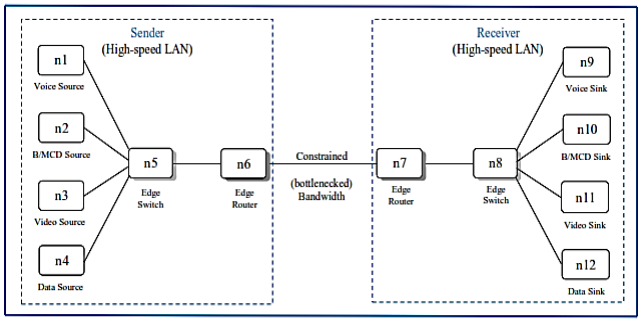

Figure 3: Simulation Network Topology

VI. Results and Discussions

The G.711 (64kbps); G.726 (32kbps) and G.729 (8kbps)) coding schemes were each used for the simulation runs. The impacts of the different coding schemes on the performance of the proposed scheduler algorithm were determined. Hence, the performance evaluation comparison of these coding schemes in the optimized QoS-based packet scheduling algorithm for bandwidth-constrained VoIP networks is presented.

A. *CODEC end-to-end Delay Determination*

Figure 4 shows a comparative analysis of the impact of the coding schemes on the performance of the proposed QoS model in terms of end-to-end packet delay. In the network, the variation of the source intensities gave 34.19%, 33.86% and 31.94% respectively for G.711, G.726 and G.729. The plot shows that the G.729 coding scheme offers the best end-to-end delay at 8kbps bit rate and therefore guarantees better network QoS performance. Recall that of the three coding schemes investigated, the G.729 has the least bit rate. This fact further validates the optimal performance of the proposed QoS model.

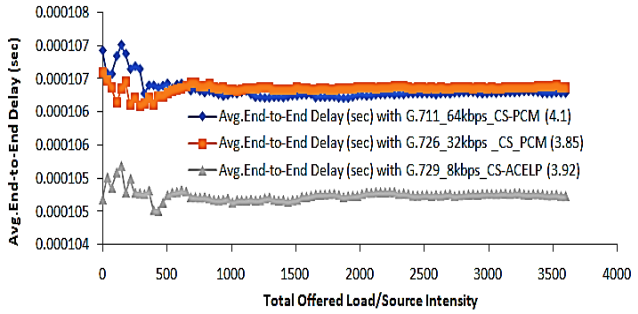

Figure 4: Average End-to-End Delay evaluation using G.711, G.726 and G,729 CODECs

B. *CODEC Packet loss Determination*

In real-time communications, packet losses occur when traffic/packets that fail to arrive the receiver timely are discarded. Eliminating or reducing packet losses therefore, optimizes the quality of service obtainable for time-sensitive services such as voice and BCMD. Figure 5 shows a comparative analysis of the impact of the coding schemes in terms of packet loss. From the Riverbed statistic engine script, the plot of G.711, 726, and 729 offered 60%, 33.33% and 6.67% respectively. The plots therefore show that G.729 guarantees lower packet loss in the network, thereby validating the optimal performance of the proposed QoS model.

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

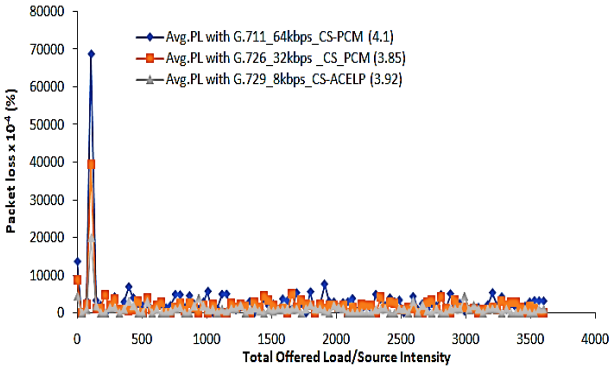

Figure 5: Average Packet Loss Probability Evaluation using G.711, G.726 and G.729 CODECs

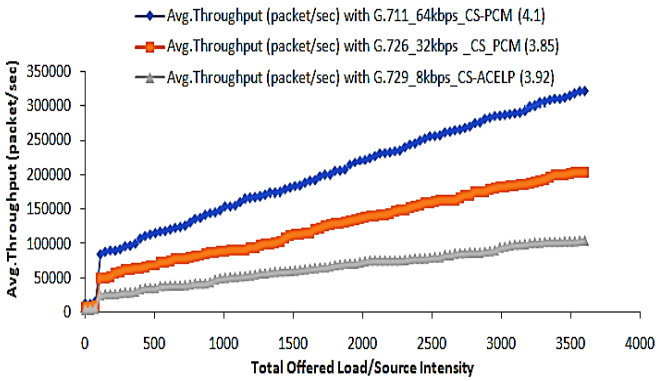

Figure 6: Average Throughput Performance Evaluation using G.711, G.726 and G.729 CODECs

C. *CODEC Throughput Determination*

It is know that the lower the CODEC bit-rate, the less bandwidth is required for transmission and this translates to efficient system performance. However, reliable delivery of speech quality is affected by the type of CODEC in place. With G.711 higher bit rates of the encoded bit-stream, this will give higher voice quality with a trade-off cost on the bandwidth or transmission capacity. The lower data rate offers good savings in transmitted data rate (i.e., lower bit rate). The implication is that this could generate delays for voice encoding. A good CODEC scheme compresses the voice signal for stable network throughput.

Figure 6 shows a comparative analysis of the impact of the coding schemes in terms of total available throughput. From the Riverbed statistic engine script, the plot of G.711, 726, and 729 offered 57.38%, 32.78% and 9.83% respectively. The plots therefore show that the G.711 coding scheme has the highest throughput, followed by G.726 and then G.729. This result is obvious since the G.711 has the highest bit rate. As such a trade-off is necessary to satisfy the demands of network QoS. For VOIP networks, G.729 offers a satisfactory throughput while maintaining optimal bandwidth consumption, packet loss probability and end-to-end, absolute transmission (mouth-to-ear) delay.

D. *CODEC Resource Utilization Determination*

Figure 7 shows the network resource utilization considering the offered load intensities. By controlling and managing network resources through priority setting for traffic sources on the network, resource utilization is fairly distributed. It was observed that with traffic connection request, resource allocation/utilization for G.711, G.726 and G.729 normalized with increasing source intensity. The utilization offerings were obtained as 44.44%, 37.04% and 18.52% respectively. This shows that G.729 had the least resource utilization, hence suitable for high performance network. This fact also validates the optimal performance of the proposed hybrid architecture since G729 has the least bit rate compared with G.711 and G.726 schemes.

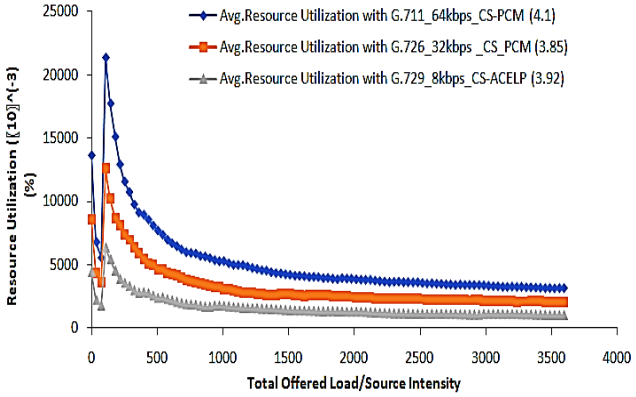

Figure 7: Average Resource Utilization Performance Evaluation using G.711, G.726 and G.729 CODECs

Table 1: Comparison of QoS Performance of the Proposed Algorithm using different CODEC Schemes

S/N	Quality of Service Performance Metrics	CODEC Schemes		
		G.711 (64kbps)	G.726 (32kbps)	G.729 (8kbps)
1	End-to-End Delay	34.19%	33.86%	31.94%
2	Packet Loss	60%	33.33%	6.67%
3	Throughput	57.38%	32.78%	9.83%
4	Resource Utilization	44.44%	37.04%	18.52%
5	Mean Opinion Score (MOS)	4.1	3.85	3.92

VII. CONCLUSIONS

A summary of the comparison of the QoS performance evaluation for G.711, 726, and 729 coding schemes is given in Table 1. The Mean opinion score (MOS) are respectively placed at 4.1, 3.85 and 3.92. From industry perspective, the highest MOS score is 4.1. But this has serious cost implications in bandwidth-constrained VoIP networks as observed in network throughput performance. Hence, the G.729 coding scheme, which has the lowest bit rate offers the

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

best performance in end-to-end delay, packet loss and resource utilization. It therefore guarantees better network QoS performance. This fact validates the optimal performance of the proposed hybrid QoS-based multimedia model in bandwidth-constrained VoIP networks. Owing to the growing volume of critical data generated by cyber-physical computer and other online electronic systems, the need for the development of more robust multimedia architectures has become imperative.

REFERENCES

[1] W. C. Chu, "Speech Coding Algorithms-Foundation and Evolution of Standardized Coders", Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

[2] S. Jelassi, G. Rubino, H. Melvin, H. Youssef and G. Pujolle, "Quality of Experience of VoIP Service: A Survey of Assessment Approaches and Open Issues", IEEE Communications Survey & Tutorials, Vol.14, No. 2, pp 491-513, Second Quarter 2012.

[3] T. Ogunfunmi and M. J. Narasimba, "Speech over VoIP Networks: Advanced Signal Processing and System Implementation", IEEE Circuits and Systems Magazine, Volume: 12, Issue: 2, pp 35-55, 2nd Quarter, 2012. DOI: 10.1109/MCAS.2012.2193436

[4] ITU-T Recommendation G.711,"Pulse Code Modulation (PCM) of Voice Frequencies", International Telecommunication Union Telecommunication Standardization Sector, T-REC-G.711-198811, 1988.

[5] S. Haykin, "Communication Systems", 4th Edition, John Wiley & Sons, Inc., 2001.

[6] ITU Recommendation G.726,"40, 32, 24, 16 kbps Adaptive Differential Pulse Code Modulation (ADPCM)", T-REC-G.726-199011, ITU - International Telegraph and Telephone Consultative Committee (CCITT), 1990.

[7] J. H. Chen, "High-Quality 16 kb/s Speech Coding with a One-Way Delay Less 2 ms", Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Albuquerque, NM, USA, USA, pp. 453–456, April 1990. DOI: 10.1109/ICASSP.1990.115747

[8] ITU-T Recommendation G.728,"Coding of Speech at 16 kbps using Low-Delay Code Excited Linear Prediction (LD-CELP)", T-REC-G.728-201206, International Telecommunication Union Telecommunication Standardization Sector, 2012.

[9] ITU-T Recommendation G.729,"Coding of Speech at 8 kbps using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)", T-REC-G.729-201206, International Telecommunication Union Telecommunication Standardization Sector, 2012.

[10] ITU-T Recommendation G.723.1, "Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbps", T-REC-G.723.1-200605, International Telecommunication Union Telecommunication Standardization Sector, 2006.

[11] V. K. Singh, "QoS-Based Techniques: Investigation and Optimization", International Journal of Engineering Science and Computing (IJESC), 6 (5), 2016, pp 5242-5246.

[12] N. Gupta, N. Kumar and H. Kumar, "Comparative Analysis of Voice CODECs over Different Environmental Scenarios in VoIP", 2nd International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, June 2018, pp 540-544. DOI: 10.1109/ICCONS.2018.8663241

[13] T. Chakraborty, I. T. Misra and S. K. Sanyal, "QoS Enhancement Techniques for Efficient VoIP Performance in Cognitive Radio Network", International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), 6, 2014, pp 413-426.

[14] L. A. Haibeh, N. Hakem and O. A. Safia, "Performance Evaluation of VoIP Calls over MANET for Different Voice CODECs", IEEE 7th Annual Computing and Communications Workshop and Conference (CCWC), Las Vegas, NV, USA, Jan 2017, pp 1-6. DOI: 10.1109/CCWC.2017.7868479

[15] C. Chen and C. Pan, "Dual Threshold Scheduling for VoIP Traffic on Downlink of WiMAX Networks", Journal of Computers, 9 (11), 2014, pp 2712-2719.

[16] H. P. Singh, S. Singh, J. Singh and S. A. Khan, "VoIP: State of the Art for Global Connectivity – A Critical Review", Journal of Network and Computer Applications (JNCA), 37, 2014, pp 365-379.

[17] A. Chhabra and D. Singh, "Assessment of VoIP E-Model over 802.11 Wireless Mesh Network", International Conference on Advances in Computer Engineering and Applications (ICACEA), Ghaziabad, India, 2015, pp 856-860.

[18] A. S. W. Marzuki, Y. K. Chai, H. Zen, L. L. Wee, K. Lias and D. A. A. Mat, "Performance Analysis of VoIP over 802.11b and 802.11e using Different CODECs", 10th International Symposium on Communications and Information Technologies (ISCIT), Tokyo, Japan, 2010, pp 244-248. DOI: 10.1109/ISCIT.2010.5664844

[19] M. O. Ortega, G. C. Altamirano and M. F. Abad, "Evaluation of the Voice Quality and Quality of Service in Real Calls using Different VoIP CODECs", IEEE Colombian Conference on Communications and Computing (COLCOM), Medellin, Colombia, 2018, pp 1-6. DOI: 10.1109/ColComCon.2018.8466727

[20] M. T. Meeran, P. Annus and Y. L. Moullec, "The Current State of Voice over Internet Protocol in Wireless Mesh Networks", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp 2567-2575.

[21] Y. Labyad, M. Moughit and A. Haqiq, "Performance Analysis and Comparative Study of Voice over IP using Hybrid CODEC", IEEE International Conference on Complex Systems (ICCS), Agadir, Morocco, May 2012, pp 1-6. DOI: 10.1109/ICoCS.2012.6458570

[22] M. O. Ortega, G. C. Altamirano, C. L. Barros and M. F. Abad, "Comparison Between the Real and Theoretical Values of the Technical Parameters of the VoIP CODECs", IEEE Colombian Conference on Communications and Computing (COLCOM), Barrauquilla, Colombia, June 2019, pp 1-6. DOI: 10.1109/ColComCon.2019.8809181

[23] A. M. Alsahlany, "Performance Analysis of VoIP Traffic over Integrated Wireless LAN and WAN using Different CODECs", International Journal of Wireless and Mobile Networks (IJWMN), 6 (3), 2014, pp 79-89.

[24] S. Sahabudin and M. Y. Alias, "End-to-End Delay Performance Analysis of Various CODECs on VoIP Quality of Service", IEEE 9th Malaysia International Conference on Communications (MICC), Kaula Lumpur, Malaysia, Dec 2009, pp 607-612. DOI: 10.1109/MICC.2009.5431426

IEEE Office - SEET Complex, Federal University of Technology, Owerri, Nigeria, https://aeuso.org
EISSN: 2305-0543, PISSN: 2411-6173

[25] A. M. Alsahlany and H. S. Rashid, "Audio Codecs Impact on Quality of VoIP Based on IEEE802.16e Considering Mobile IP Handover", American Journal of Networks and Communications, 4 (3), 2015, pp 59-66.

[26] H. M. T. Al-Hilfi, A. H. Najim and A. M. Alsahlany, "Evaluation of WIMAX Network Performance of Baghdad City using Different Audio CODECs", 16th RoEduNet Conference: Networking in Education and Research (RoEduNet), Targu Mures, Romania, Oct 2017, pp 1-5. DOI: 10.1109/ROEDUNET.2017.8123762

[27] A. Zmily and D. A. Saymeh, "Quality CODEC Interface for VoIP on Differentiated Services Networks", International Conference on Smart Communications in Network Technologies (SaCoNet), Paris, France, June 2013, pp 1-5. DOI: 10.1109/SaCoNet.2013.6654563

[28] S. Gurrapu, S. Mahta and S. Panbude, "Comparative Study for Performance Analysis of VoIP CODECs over WLAN in Non-Mobility Scenarios, International Journal of Information Technology, Modeling and Computing (IJITMC), 4 (4), 2016, pp 1-16.

[29] T. Uhl, "QoS by VoIP under Use Different Audio Codecs", IEEE Joint Conference – Acoustics, Ustka, Poland, 2018, pp 1-4.

[30] M. Aamir and S. M. A. Zaidi, "QoS Analysis of VoIP Traffic for Different CODECs and Frame Counts per Packet in Multimedia Environment using OPNET", 15th International Multitopic Conference (INMIC), Islamabad, Pakistan, 2012, pp 1-7. DOI: 10.1109/INMIC.2012.6511508

[31] J. N. Dike and C. I. Ani, "Design and Simulation of Voice and Critical Data Priority Queue (VCDPQ) Scheduler for Constrained-Bandwidth VoIP Networks", International Journal of Scientific and Engineering Research (IJSER), 9 (9), 2018, pp 2050-2056.

[32] J. N. Dike and C. I. Ani, "Design and Analysis of Voice and Critical Data Priority Queue (VCDPQ) Scheduler for Constrained-Bandwidth VoIP Networks", American Journal of Engineering Research (AJER), 7 (10), 2018, pp 154-167.

[33] J. N. Dike and C. I. Ani, "An Optimized Critical Data and Voice Hybrid Scheduler for Cyber-Physical Computer Systems (CPCS) in Constrained-Bandwidth VoIP Networks", 1st International Conference on Mechatronics, Automation and Cyber-Physical Computer Systems, (MAC-2019), Federal University of Technology, Owerri (FUTO), Nigeria, 2019, pp 180-186.

[34] J. N. Dike, "Optimizing the Quality of Service of Constrained-Bandwidth Voice over Internet Protocol Networks", Ph.D Thesis, Department of Electronic Engineering, University of Nigeria, Nsukka, 2019.

[35] J. N. Dike and C. I. Ani, "Validation of Critical Data and Voice Hybrid Scheduler (CDVHS) for Cyber-Physical Computer Systems (CPCS) in Constrained-Bandwidth VoIP Networks", International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), 10 (37), 2020, pp 4673-4683.

[36] J. N. Dike and C. I. Ani, "Performance Evaluation of an Optimized Hybrid Packet Scheduler for Bandwidth-Constrained Voice over Internet Protocol Networks", Uniport Journal of Engineering and Scientific Research (UJESR), 5, Special Issue, 2020, pp 129-139.

[37] J. D. C. Little and S. C. Graves, "Little's Law", D. Chhajed and T. J. Lowe (eds.) Building Intuition: Insights From Basic Operations Management Models and Principles, © Springer Science + Business Media, LLC 2008, DOI: 10.1007/978-0-387-73699-0.

[38] A. Leon - Garcia, 'Probability and Random Processes for Electrical Engineering' (2nd Edition), Addison Wesley 1993.

[39] A. Leon-Garcia and I. Widjaja, "Communication Networks: Fundamental Concepts and Key Architectures", McGraw-Hill Companies, Inc. New York, 2000.

[40] J. K. Sharma, "Operations Research: Theory and Applications", Macmillan Publishers India Ltd., 5th Ed., 2013.

[41] V. Jacobson, K. Nichols and K. Poduri, "Expedited Forwarding PHB Group", Network Working Group RFC 2598, Internet Engineering Task Force (IETF), 1999.

[42] J. Heinanen, F. Baker, W. Weiss and J. Wroclawski, "Assured Forwarding PHB Group", Network Working Group RFC 2597, Internet Engineering Task Force (IETF), June 1999.

[43] S. G. Chaudhuri, C. S. Kumar and R. V. RajaKumar, "Validation of a DiffServ based QoS Model Implementation for Real-Time Traffic in a Test Bed", Procedings of the IEEE National Conference on Communications (NCC), Kharagpur, India, 2012. DOI: 10.1109/NCC.2012.6176841

[44] S. Gordon, "Introduction to OPNET Modeler", Sirindhorn International Institute of Technology, Thammasat University, June 2010.

[45] Riverbed Technology, 'Riverbed Modeler: A Suite of Protocols and Technologies with a Sophisticated Development Environment' Riverbed Technology, San Francisco, USA, 2018.