American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-7, Issue-10, pp-154-167
www.ajer.org
Research Paper Open Access

Design And Analysis Of Voice And Critical Data Priority Queue (VCDPQ) Scheduler For Constrained-Bandwidth Voip Networks

J. N. Dike[1] And C. I. Ani[2]
[1](Department of Electrical/Electronic Engineering, University of Port Harcourt, Nigeria)
[2](Department of Electronic Engineering, University of Nigeria, Nsukka, Nigeria)
Corresponding Author: J. N. Dike

ABSTRACT : *Voice over Internet Protocol (VoIP) networks are converged information superhighway systems for transporting multimedia traffics, such as: voice, video and data traffics. In the quest to optimize the Quality of Service (QoS) of these networks, many schemes have been proposed to address the needs of both real-time and non real-time traffic flows. In most of the schemes, low latency and packet loss to the streaming flows as well as fair resource utilization are guaranteed only if the rate of the real-time traffic flow is a small fraction of the link capacity. Again, earlier solutions to the QoS challenges of VoIP networks have been focused on giving precedence to voice traffic at the expense of time-sensitive business/mission critical data (B/MCD) traffics. This work therefore investigates the contributions of delay and packet loss impairment factors to the overall quality degradation of VoIP networks. It proposes an optimized voice and B/MCD fair and priority traffic scheduler for constrained-bandwidth VoIP networks. The scheduler incorporates mechanisms to achieve a graceful trade-off between priority and fairness to all traffics. Riverbed (OPNET) Modeler was used to validate the proposed architecture. Simulation results obtained show that the proposed architecture guarantees good mouth-to-ear delay and packet loss probability.*
KEYWORDS - *Critical data, latency, packet loss, multimedia, quality of service.*

DATE OF SUBMISSION:04-10-2018 DATE OF ACCEPTANCE: 16-10-2018

I. INTRODUCTION

Voice over Internet Protocol (VoIP) networks are emerging converged information super-highway systems for transporting multimedia traffics (comprising voice, video and data traffics). These traffics are broadly classified as real-time and non real-time. Real-time traffics (such as voice) as well as business/mission critical data ((B/MCD), such as real-time online purchases, security alerts, bank transfers, weather forecasts, remote/emergency environmental monitoring, disaster alerts, military commands, remote industrial control systems, and so on) are delay sensitive and require small but assured amount of bandwidth, low delay, low jitter and low packet loss. Data applications (non real-time traffics) such as e-mail, World Wide Web (WWW), File Transfer Protocol (FTP), and so on, need more bandwidth, but can tolerate the delay and jitter impairments [1, 2]. By segregating network traffic into different classes and applying different forwarding treatments, a wide range of throughput, loss and delay performance can be achieved [3]. The major quality of service (QoS) impairment factors in VoIP networks include bandwidth limitation, delay (latency), jitter (delay variation) and packet/frame loss [4, 5, 6]. These factors are typical performance metrics of computer networks and are used in defining QoS.

This work addresses the effect of delay and packet/frame loss on the perceived voice quality in constrained-bandwidth VoIP networks. The dominant causes of delay in packet networks are fixed propagation delays on wide area links and variable queuing delays in switches and routers. Since propagation delays are a fixed property of the network topology, overall delay and jitter are minimized when the variable queuing delays are minimized. Furthermore, if queues remain short relative to the buffer space available, packet loss is also kept to a minimum.

Owing to the fact that no single scheme can effectively solve these transmission (voice quality) impairment factors [7], a hybrid scheme is hereby proposed. The proposal models an approach of adaptively evaluating and policing incoming Internet Protocol (IP) flows as well as classifying and mapping different

traffic types for individual applications or users. The architecture incorporates mechanisms to achieve a graceful tradeoff between priority and fairness to all traffics as a solution to the transmission impairments in the evolving growing VoIP networks. An analytical appraisal of the developed architecture is hereby presented.

A critical optimal network QoS requirement demands that the offered load ($\rho_{net} = \beta_{net}/\mu_{net}$) 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. The analysis ensures adequate bounding in the inbuilt Congestion Control mechanisms of the proposed model. 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 γ bits/second in the Token Bucket module while the shared transmission link services the flow regulated by the weighted round robin (WRR) scheduler from the Differentiated Services (DiffServ) module. This implies that the average delay and packet loss encountered by traffic flows in each domain are accounted for by these links. Riverbed (OPNET) modeler [8] was used to simulate the performance of the proposed scheme. To implement the architecture, an Internet Service Provider's (ISP's) network is configured accordingly.

II. EARLIER PROPOSALS

Earlier related works in optimizing the QoS of a VoIP network have been focused on traffic-scheduling algorithms to ensure either minimum traffic delay constraints or fair resource sharing to all applications running on the network [9]. A QoS-guaranteed network normally differentiates between different types of traffic and provides different treatments to the traffics. This is made possible by using either the type-of-service (ToS) [10] bits or the differentiated services (DiffServ) [11, 12, 13, 14, 15, 16] field in the IP header, or still through the use of signalling protocols such as: resource reservation protocol (RSVP) [17, 18] and multi-protocol label switching (MPLS) [19]. Traffic identification can also be implemented by configuring network devices to support prioritization based on physical port, protocol, IP address, transport address or packet length [20]. Several queuing strategies have been proposed as solution to support the delay-constrained traffics in a best-effort network [21, 22, 23, 24, 25]. In all these proposals however, no precedence was given to B/MCD traffic flows.

III. METHODOLOGY

3.1 Design of an Optimized QoS Model

The proposed optimized QoS architecture is an integration of several technologies. It is comprised of the Packet Classifier, the Token Bucket, the Differentiated Services (DiffServ) and the Weighted Round Robin (WRR) Scheduler modules. The hybrid architecture [9] is illustrated in Figure 1.

The Packet Classifier module consists of two packet classifiers. Classifier1 is used to classify the packets of the incoming source traffic (p) into two main classes, namely: voice (p1) and non-voice (p2) flows [26, 27]. Packet Classifier2 is used to classify 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 business/mission-critical data flows the necessary priority and fairness they deserve.

The dynamics of network traffic flow and packet distinguishing is implemented by the input service routine architecture, which applies traffic congestion avoidance controls [28, 29] to the incoming flows and places incoming packets into separate queues for subsequent processing by inspecting the type-of-service (ToS) [10] bits in the packet IP header. 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.

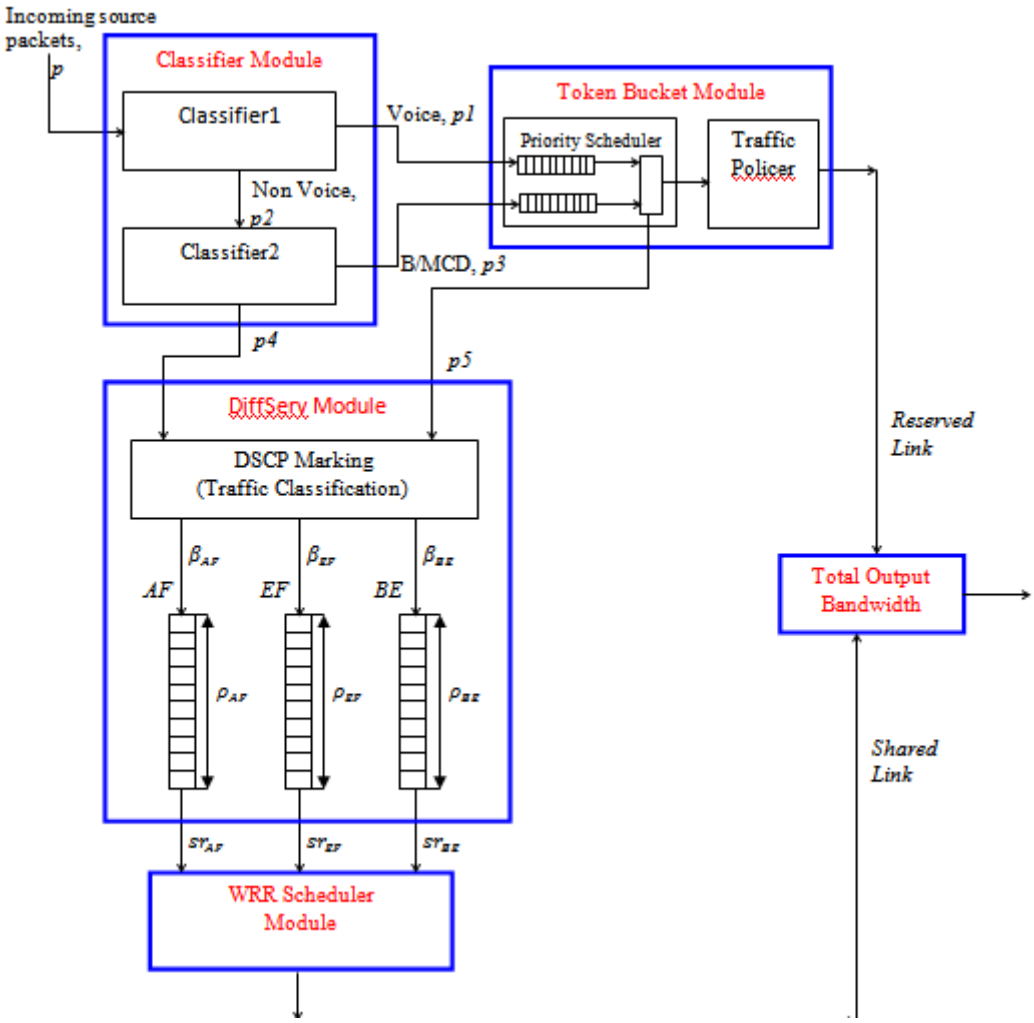


Figure 1: The optimized hybrid scheduler architecture

Legend

β_AF, β_EF and β_BE = average input packet rate of AF, EF and BE traffics.

ρ_AF, ρ_EF and ρ_BE = buffer size (or average queue length) for AF, EF and BE traffics.

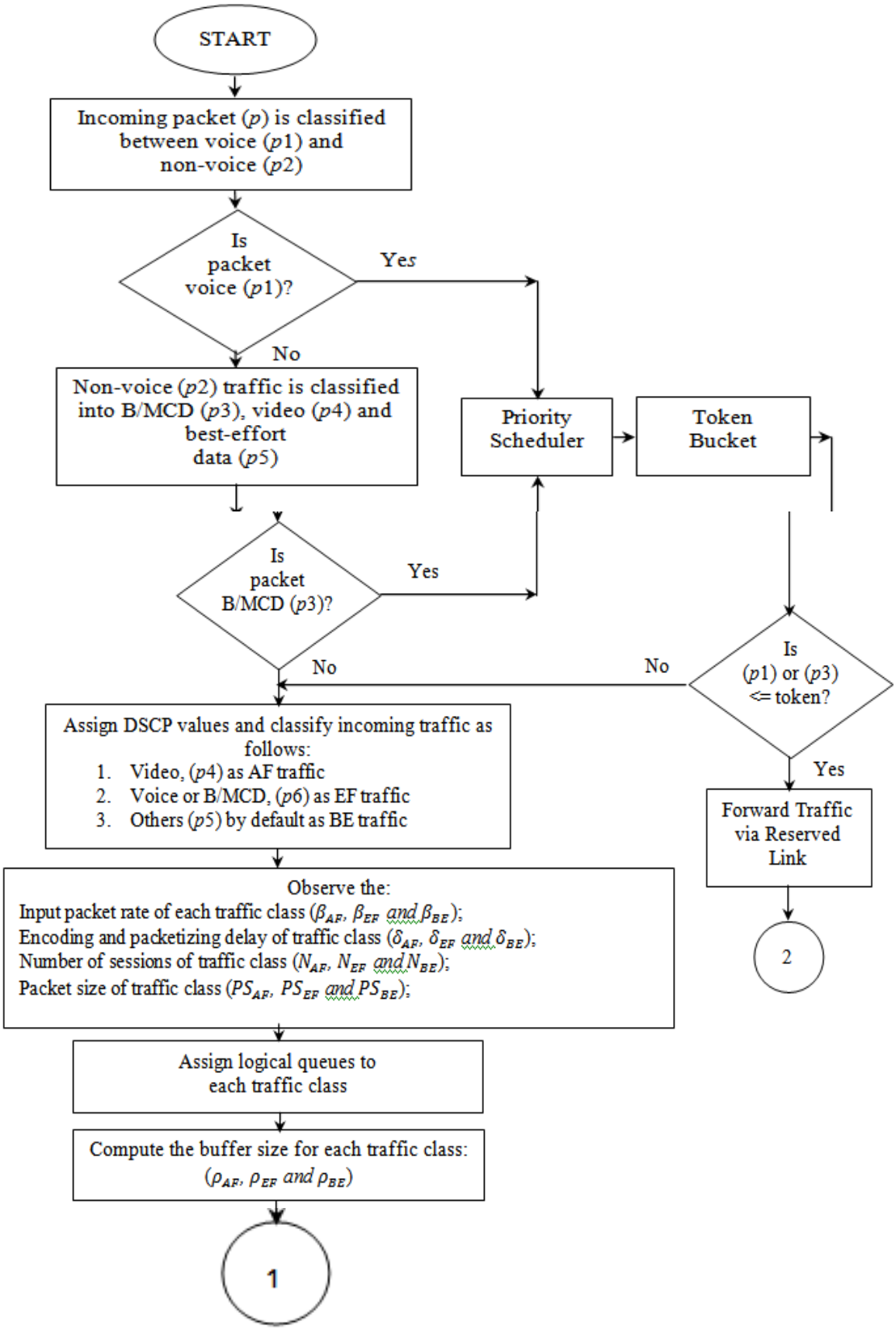
〖sr〗_AF, 〖sr〗_EF and 〖sr〗_BE = service rate (or queue weight) for AF, EF and BE traffics.

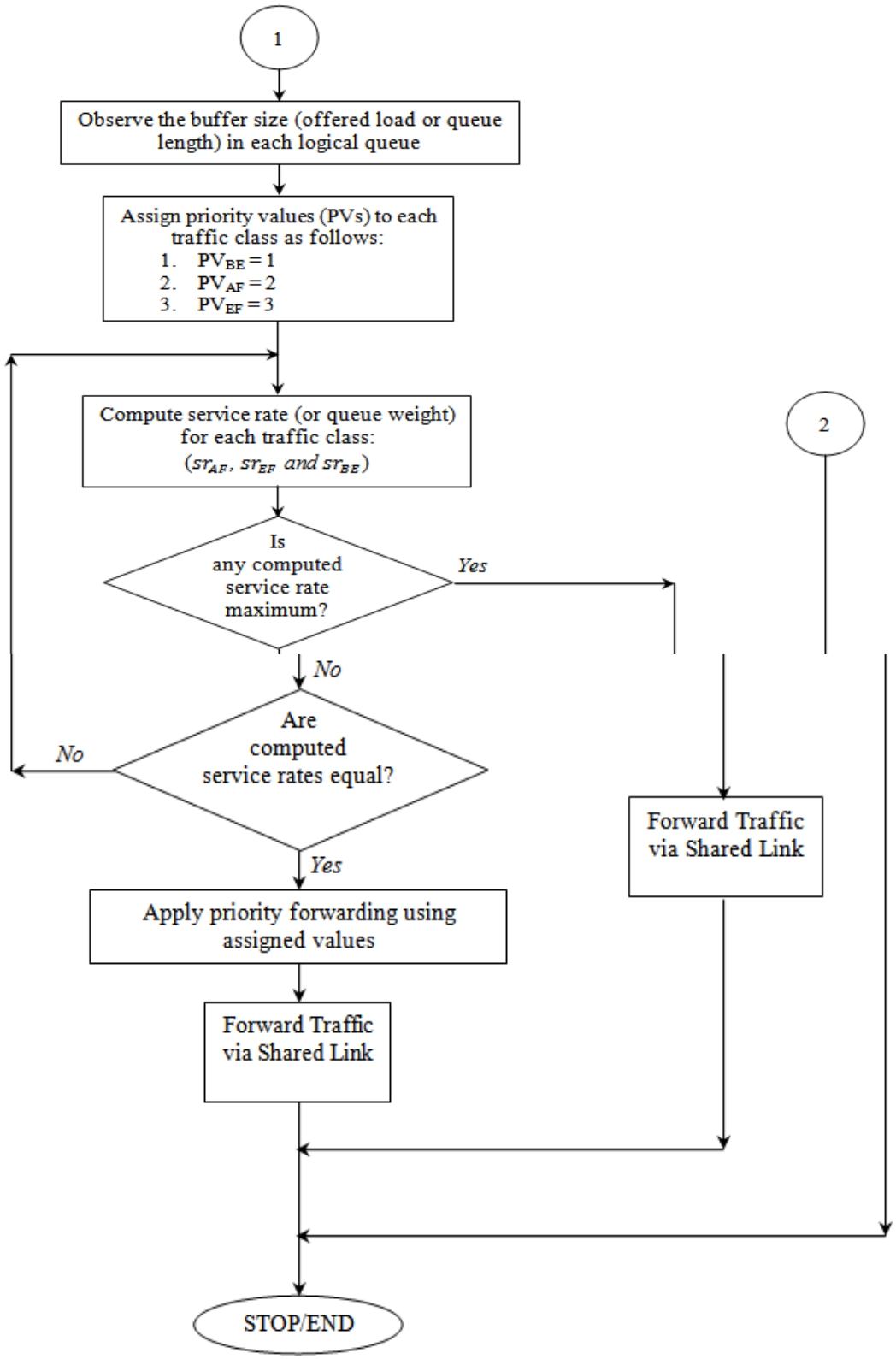
The Token Bucket module is used to split the incoming voice or business/mission-critical data traffic into two sub-flows [26]. The first sub-flow is a well shaped flow with maximum rate equal to γ bits/second generated by the Token Bucket. The second sub-flow is the packet (p5 - still of voice or business/mission-critical data traffic) rejected by the Token Bucket.

In the DiffServ module, video traffic is mapped to Assured Forwarding (AF) traffic class. Voice or business/mission-critical data 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.

A structured signal flowchart describing the proposed scheduling algorithm is presented in Figure 2. The top-down design approach was used to define the various activities performed at every level of abstraction (module).


START
Incoming packet (p) is classified between voice (p1) and non-voice (p2)
Is packet voice (p1)?
Yes
No
Non-voice (p2) traffic is classified into B/MCD (p3), video (p4) and best-effort data (p5)
Priority Scheduler
Token Bucket
Is packet B/MCD (p3)?
Yes
No
Is (p1) or (p3) <= token?
No
Yes
Assign DSCP values and classify incoming traffic as follows:
1. Video, (p4) as AF traffic
2. Voice or B/MCD, (p6) as EF traffic
3. Others (p5) by default as BE traffic
Forward Traffic via Reserved Link
2
Observe the:
Input packet rate of each traffic class (β_{AF}, β_{EF} and β_{BE});
Encoding and packetizing delay of traffic class (δ_{AF}, δ_{EF} and δ_{BE});
Number of sessions of traffic class (N_{AF}, N_{EF} and N_{BE});
Packet size of traffic class (PS_{AF}, PS_{EF} and PS_{BE});
Assign logical queues to each traffic class
Compute the buffer size for each traffic class: (ρ_{AF}, ρ_{EF} and ρ_{BE})
1


1
Observe the buffer size (offered load or queue length) in each logical queue
Assign priority values (PVs) to each traffic class as follows:
1. $PV_{BE} = 1$
2. $PV_{AF} = 2$
3. $PV_{EF} = 3$
Compute service rate (or queue weight) for each traffic class: (sr_{AF}, sr_{EF} and sr_{BE})
Is any computed service rate maximum?
Yes
No
Are computed service rates equal?
No
Yes
Apply priority forwarding using assigned values
Forward Traffic via Shared Link
2
Forward Traffic via Shared Link
STOP/END

3.2 An Analytical Analysis of the Developed Architecture

The approximation method of analyzing queuing networks by decomposition [30] is adopted in this analysis. This strategy is considered as decomposition of the whole network or as aggregation of portions of the network. In the method, the arrival rate is assumed to be Poisson and the service times of network elements are exponentially distributed. A summary of analysis by decomposition is given in the following steps [31]:

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

3.2.1 In The Token Bucket Module:

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 $\gamma\ bits/second$ and is expressed as (αTP), where α is the splitting ratio of the Token Bucket and is defined as ($0 \leq \alpha \leq 1$) [26]. 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 1 implies that $TP_{rej} = 0$. These imply that:

$$\left.\begin{aligned} TP_{rev}\ (TP_{V-rev}\ or\ TP_{B/MCD-rev}) &= \gamma = (\alpha TP) \\ TP_{rej}\ (TP_{V-rej}\ or\ TP_{B/MCD-rej}) &= [(1-\alpha)TP] \end{aligned}\right\} \quad \dots \quad (1)$$

Let us now consider a typical run of the proposed hybrid scheduling model during a time interval (t_1, t_2) 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 y bits/second. Hence, the upper bound of bits serviced by this flow in the time interval (t_1, t_2) is equal to $\gamma(t_1, t_2)$. 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, even with the expected increase in the voice traffic.

Assuming that the Token Bucket is empty (of tokens) at $t_1 = 0$ and full at t_2, it follows that the total number of packets arriving the Bucket at time t_2 is defined by [32]:

$$N_{TB}(t) = TP(t) - \alpha TP(t) - [(1-\alpha)TP](t) \quad \dots \quad \dots \quad (2)$$

This implies that up until time t_2, the number of packets that has entered the bucket is $TP(t) - [(1-\alpha)TP](t)$ and that $\alpha TP(t)$ of these packets has been transmitted via the dedicated link by time t_2. The long-run total packet arrival rate at the bucket is defined by:

$$\beta_{TB} = \frac{TP(t)}{t}\ packets/second \quad \dots \quad \dots \quad (3)$$

The average number of packets transmitted through the reserved link (which is the throughput of the Token Bucket) is equal to the long-run transmission rate and is defined by:

$$Throughput_{TB} = \frac{\alpha TP(t)}{t}\ packets/second \quad \dots \quad \dots \quad (4)$$

The fraction of arriving packets that are rejected from the bucket is therefore defined by:

$$P_{rej} = \frac{[(1-\alpha)TP](t)}{TP(t)} \quad \dots \quad \dots \quad \dots \quad \dots \quad (5)$$

Assuming the bucket to be empty at t_1, the average number of packets is defined by:

$$Av[N_{TB}] = \frac{1}{t_2}\int_{t_1}^{t_2} \alpha TP(t)dt\ packets \quad \dots \quad \dots \quad \dots \quad (6)$$

Applying Little's Theorem [33, 32] in the Token Bucket gives:

$$\left.\begin{aligned} Av[N_{TB}] &= \beta_{TB}(1 - P_{pl-TB})Av[D_{TB}] \\ Av[D_{TB}] &= \frac{1}{\beta_{TB}(1-P_{pl-TB})}Av[N_{TB}] \end{aligned}\right\} \quad \dots \quad \dots \quad (7)$$

$\gg\gg$

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 (t_1, t_2), 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. The average number of packets in the bucket is given by:

$$\boldsymbol{Av[N_{TB}]} = \left\{\left[\frac{\rho_{TB}}{(1-\rho_{TB})}\right] - \left[\frac{(K_{TB}+1)\rho_{TB}{}^{K_{TB}+1}}{(1-\rho_{TB}{}^{K_{TB}+1})}\right]\right\} \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

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 is defined as [32]:

$$\boldsymbol{P_{pl-RL}} = \boldsymbol{P_{pl-TB}} = \boldsymbol{\rho_{TB}^{K_{TB}}}\left[\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right] \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

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

$$\boldsymbol{Av[D_{RL}]} = \boldsymbol{Av[D_{TB}]} = \frac{\boldsymbol{Av[N_{TB}]}}{\boldsymbol{\beta_{TB}(1-P_{pl-TB})}} \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

$$= \left[\frac{\left\{\left[\frac{\rho_{TB}}{(1-\rho_{TB})}\right] - \left[\frac{(K_{TB}+1)\rho_{TB}{}^{K_{TB}+1}}{(1-\rho_{TB}{}^{K_{TB}+1})}\right]\right\}}{\left\{\rho_{TB}\mu_{TB}\left[1-\rho_{TB}^{K_{TB}}\left(\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right)\right]\right\}}\right] \quad \ldots \quad \ldots \quad (11)$$

The probability of the average time spent waiting in queue before servicing is therefore given by:

$$\boldsymbol{Av[W_{T-RL}]} = \boldsymbol{Av[D_{RL}]} - \boldsymbol{Av[S_{T-RL}]}$$

$$= \left[\left(\frac{\left\{\left[\frac{\rho_{TB}}{(1-\rho_{TB})}\right] - \left[\frac{(K_{TB}+1)\rho_{TB}{}^{K_{TB}+1}}{(1-\rho_{TB}{}^{K_{TB}+1})}\right]\right\}}{\left\{\rho_{TB}\mu_{TB}\left[1-\rho_{TB}^{K_{TB}}\left(\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right)\right]\right\}}\right) - \left(\frac{1}{\mu_{TB}}\right)\right] \quad \ldots \quad \ldots \quad (12)$$

where $Av[S_{T-RL}] = 1/\mu_{TB}$ is the average packet servicing (or transmission) time.

3.2.2 In the DiffServ Module

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 [34]. Similarly, video traffic is marked and classified by Assured Forwarding (AF) DSCP value [35] 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 [24]:

$$\boldsymbol{\rho_{EF}} = \frac{\boldsymbol{\beta_{EF}} \times \boldsymbol{\delta_{EF}} \times \boldsymbol{N_{EF}}}{\boldsymbol{PS_{EF}}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (13)$$

$$\boldsymbol{\rho_{AF}} = \frac{\boldsymbol{\beta_{AF}} \times \boldsymbol{\delta_{AF}} \times \boldsymbol{N_{AF}}}{\boldsymbol{PS_{AF}}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (14)$$

$$\boldsymbol{\rho_{BE}} = \frac{\boldsymbol{\beta_{BE}} \times \boldsymbol{\delta_{BE}} \times \boldsymbol{N_{BE}}}{\boldsymbol{PS_{BE}}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (15)$$

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.

The input packet arrival rate (β) is collected by monitoring/observing the characteristics of the incoming service traffic, the allowable packet delay (δ) is obtained from the QoS requirements of the traffic while the number of sessions (N) is given by the admission control mechanism. The number of sessions is a function of the following:

(i) total available shared link bandwidth (SLB) regulated by the WRR scheduler,
(ii) the total buffer (TB) available at the router,

(iii) the average input packet arrival rate, and
(iv) the average packet size (PS).

3.2.3 In the WRR Scheduler Module

In the WRR scheduler module, 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 (t_1, t_2) 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 [24] for the three classes of traffic are given as follows:

$$\boldsymbol{sr_{EF}(t)} = \frac{\rho_{EF}(t) \times PV_{EF}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)} \quad \ldots \quad \ldots \quad (16)$$

$$\boldsymbol{sr_{AF}(t)} = \frac{\rho_{AF}(t) \times PV_{AF}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)} \quad \ldots \quad \ldots \quad \ldots \quad (17)$$

$$\boldsymbol{sr_{BE}(t)} = \frac{\rho_{BE}(t) \times PV_{BE}}{\rho_{EF}(t) + \rho_{AF}(t) + \rho_{BE}(t)} \quad \ldots \quad \ldots \quad (18)$$

where: PV_{EF}, PV_{AF} and PV_{BE} are respectively the priority values for EF, AF and BE traffics at the interval (t_1, t_2).

Equations (16-18) clearly show 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 (t_1, t_2). This implies that the service queue that has the longest queue length is serviced 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. 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 architecture forwards the available service queue, and so on.

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

$$\boldsymbol{Throughput_{EF}} = \frac{(R_{T-SL}) \times (sr_{EF})}{sr_{WRR}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (19)$$

$$\boldsymbol{Throughput_{AF}} = \frac{(R_{T-SL}) \times (sr_{AF})}{sr_{WRR}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (20)$$

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

where:

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

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

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 load offered by EF, which is the highest priority class is given as:

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

Similarly, the load offered by AF, which is the next highest priority class is given as:

$$\boldsymbol{\rho_{AF} = \beta_{AF} Av[S_{AF}]} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (24)$$

while the load offered by BE, which is the least priority class is given as:

$$\boldsymbol{\rho_{BE} = \beta_{BE} Av[S_{BE}]}. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (25)$$

In a typical priority class-based system [32], if the total offered load for C priority classes is less than 1, that is:

$$\boldsymbol{\rho = \rho_1 + \rho_2 + \cdots + \rho_C < 1} \quad \ldots \quad \ldots \quad \ldots \quad (26)$$

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

$$\boldsymbol{Av[W_C]} = \frac{\beta Av[S^2]}{(1 - \rho_1 - \rho_2 - \ldots - \rho_{C-1})(1 - \rho_1 - \rho_2 - \ldots - \rho_C)} \quad \ldots \quad (27)$$

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:

$$\boldsymbol{\rho_{WRR} = \rho_{EF} + \rho_{AF} + \rho_{BE} < 1} \quad \ldots \quad \ldots \quad \ldots \quad (28)$$

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

$$Av[W_{EF}] = \frac{\beta_{WRR}\, Av[S^2_{WRR}]}{(1-\rho_{EF})} \quad \ldots \quad \ldots \quad \ldots \quad (29)$$

$$Av[W_{AF}] = \frac{\beta_{WRR}\, Av[S^2_{WRR}]}{(1-\rho_{EF})(1-\rho_{EF}-\rho_{AF})} \quad \ldots \quad \ldots \quad (30)$$

$$Av[W_{BE}] = \frac{\beta_{WRR}\, Av[S^2_{WRR}]}{(1-\rho_{EF}-\rho_{AF})(1-\rho_{EF}-\rho_{AF}-\rho_{BE})} \quad \ldots \quad (31)$$

where:

$$Av[S^2_{WRR}] = \frac{\beta_{EF}}{\beta_{WRR}} Av[S^2_{EF}] + \frac{\beta_{AF}}{\beta_{WRR}} Av[S^2_{AF}] + \frac{\beta_{BE}}{\beta_{WRR}} Av[S^2_{BE}] \quad \ldots \quad \ldots \quad (32)$$

and $$\beta_{WRR} = \beta_{EF} + \beta_{AF} + \beta_{BE} \quad \ldots \quad \ldots \quad \ldots \quad (33)$$

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:

$$Av[D_{EF}] = Av[W_{EF}] + Av[S_{EF}] \quad \ldots \quad \ldots \quad (34)$$

$$Av[D_{AF}] = Av[W_{AF}] + Av[S_{AF}] \quad \ldots \quad \ldots \quad (35)$$

$$Av[D_{BE}] = Av[W_{BE}] + Av[S_{BE}] \quad \ldots \quad \ldots \quad (36)$$

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:

$$Av[D_{SL}] = Av[D_{WRR}] = \frac{1}{3}\{Av[D_{EF}] + Av[D_{AF}] + Av[D_{BE}]\} \quad (37)$$

The packet-loss probability of the shared link is defined as [36]:

$$P_{pl-SL} = P_{pl-WRR} = \rho_{WRR}^{K_{WRR}} \left[\frac{(1-\rho_{WRR})}{\left(1-\rho_{WRR}^{(K_{WRR}+1)}\right)}\right] \quad \ldots \quad \ldots \quad \ldots \quad (38)$$

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

3.2.4 The Average Total Delay and Packet-loss Probability of the Developed Model

By aggregating the mean delays of the reserved and shared links, the total average delay of the proposed optimized model $Av[D_{MOD}]$ is obtained from the expression:

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

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

where $(\beta_{net}\,(1 - P_{pl-net}))$ is the actual packet arrival rate to the network while $(\beta_k (1 - P_{pl-k}))$ is the actual packet arrival rate to, and (D_k) is the delay in, every transmission link in every domain constituting the network.

$$\gg Av[D_{MOD}] = \left[\frac{1}{\beta_{MOD}(1 - P_{pl-MOD})}\{\beta_{TB}(1 - P_{pl-TB}) Av[D_{TB}] + \beta_{WRR}(1 - P_{pl-WRR}) Av[D_{WRR}]\}\right] \quad \ldots \quad \ldots \quad \ldots \quad (39)$$

where $\beta_{MOD}(1 - P_{pl-MOD})$ is the total actual packet arrival rate to the optimized model; $\beta_{TB}(1 - Ppl-TB$ and DTB are respectively the actual packet arrival rate and delay for the reserved link while $\beta WRR1 - Ppl-WRR$ and $DWRR$ are respectively 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 from equations (9 and 38) as:

$$P_{pl-MOD} = P_{pl-RL} + P_{pl-SL}$$

$$= \left[\left[\rho_{TB}^{K_{TB}} \left(\frac{(1-\rho_{TB})}{\left(1-\rho_{TB}^{(K_{TB}+1)}\right)}\right)\right] + \left[\rho_{WRR}^{K_{WRR}} \left(\frac{(1-\rho_{WRR})}{\left(1-\rho_{WRR}^{(K_{WRR}+1)}\right)}\right)\right]\right] \quad \ldots \quad \ldots \quad \ldots \quad (40)52)$$

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 (equations 19-21) show that only a fraction of the capacity of the link is allocated to each aggregated flow.

IV. SIMULATION TOPOLOGY

The simulation of the voice and critical data priority queue (CCDPQ) scheduler for constrained-bandwidth VoIP networks is implemented using Riverbed (OPNET) Modeler v17.5. Riverbed Technology Inc is a leader in Application Performance Infrastructure that delivers the most complete platform for Location-Independent Computing. Riverbed Modeler provides a modelling and simulation environment for designing communication protocols and network equipment. It models and analyses the behavior of the entire network, including its routers, switches, protocols and servers as well as predicts the performance of IT infrastructures including individual applications and networking technologies [8].

The simulation network topology is shown in Figure 3. Nodes n1, n2, n3 and n4 are respectively the voice, business/mission-critical data (B/MCD), video and best-effort data sources (or generators). At the sending end, these nodes are connected to the edge switch (n5) via 100Mbps links. The switch is connected to the edge router (n6) via a 1Gbps link. The edge router is then connected to the network via a constrained- (bottlenecked-) or low-bandwidth link. The connection is similar but reversed at the receiving end. The marking of packets is performed by the edge switch (N5) while scheduling of packets through the network is performed by the edge router (N6). The local area networks at the sender and receiver ends are high-speed LANs. In this paper, the simulation focuses on end-to-end or mouth-to-ear (M2E) delay and packet loss probability achieved by the proposed voice and critical data priority queue (CCDPQ) scheduler.

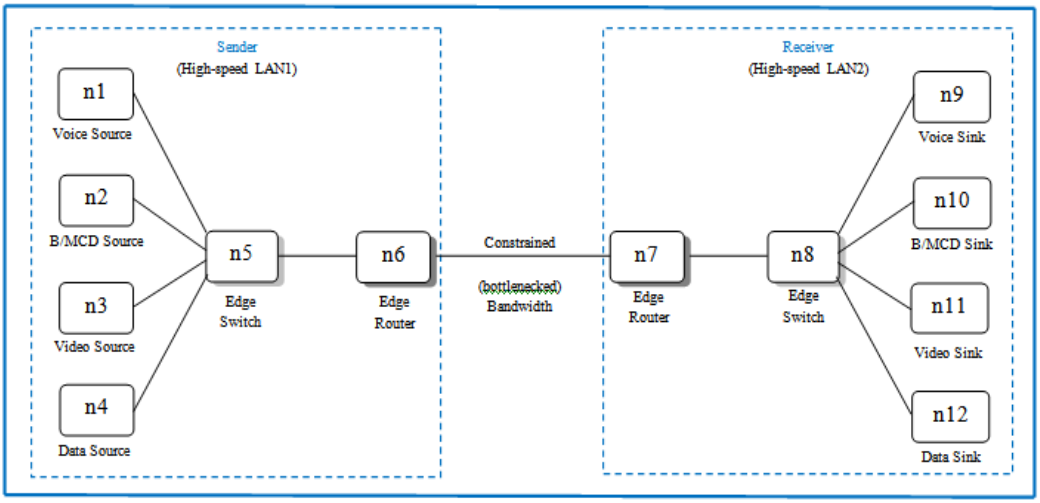


Figure 3: Simulation Network Topology

V. SIMULATION RESULTS AND DISCUSSIONS

Packet delay and packet loss QoS performance statistics for each traffic were monitored at the receiver end by increasing the load in the link with the proposed QoS model enabled and with the model disabled. From the performance statistics at the receiver end, it was observed that enabling the proposed QoS model improves the overall performance of the network.

Figure 4 shows packet queuing (end-to-end or mouth-to-ear (M2E)) delays with and without the proposed algorithm enabled plotted against time, when 70% (of total source intensity) input data rates of voice and B/MCD traffics are transmitted with 30% (of total source intensity) input data rates of video and best-effort data over the network during one simulation run from one domain in the network. Packet delay increases abruptly with time when the QoS model is disabled and gradually within acceptable limits when the model is enabled. The model therefore guarantees optimized QoS for voice and B/MCD traffics in terms of end-to-end packet delay.

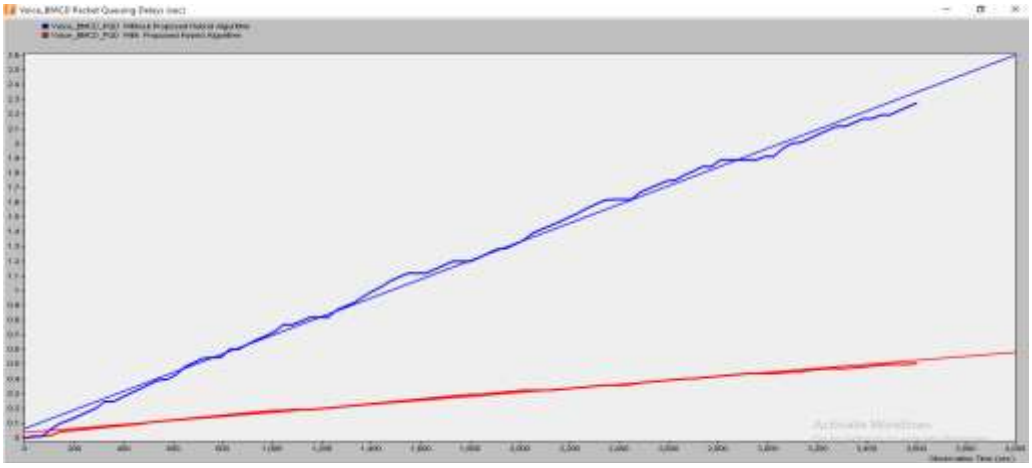


Figure 4: Network Packet Queuing Delay with and without enabling Proposed Algorithm

Figure 5 shows the validation plot of packet queuing delay with and without the proposed hybrid algorithm enabled against total source intensity for all simulation runs. Using the G729 CODEC scheme, the packet queuing delay was reduced to 17.24% while without the algorithm; it was defaulted at 82.76%. This means that traffic provisioning will be reasonably fast under the constrained-bandwidth scenario without much queuing delay experienced by queues running on the network than when the algorithm is disabled. This is very significant for the constrained-bandwidth VoIP network running time-sensitive services.

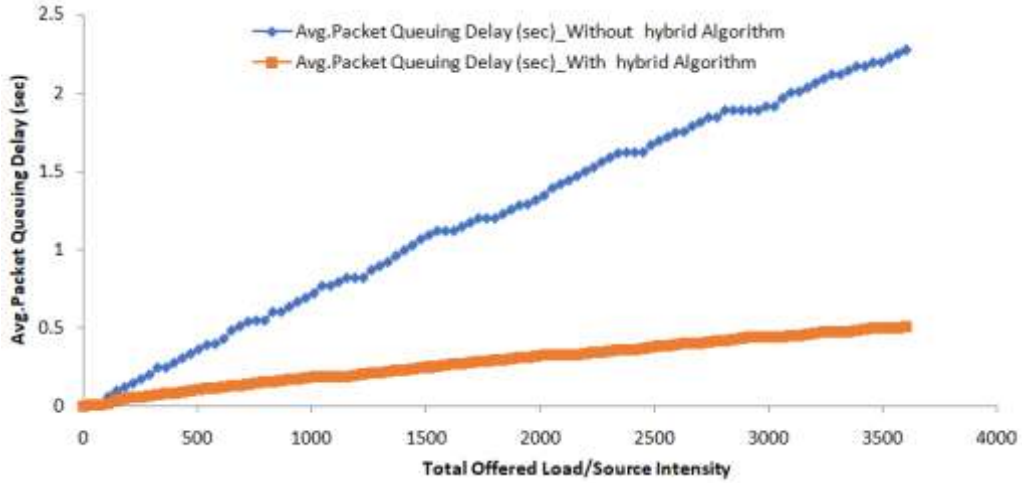


Figure 5: Validation plot of Packet Queuing Delay against Total Source Intensity with and without the proposed Hybrid Algorithm.

Figure 6 shows the packet loss probability with and without the proposed algorithm enabled plotted against time, when 70% (of total source intensity) input data rates of voice and B/MCD traffics are transmitted with 30% (of total source intensity) input data rates of video and best-effort data over the network during one simulation run from one domain in the network.

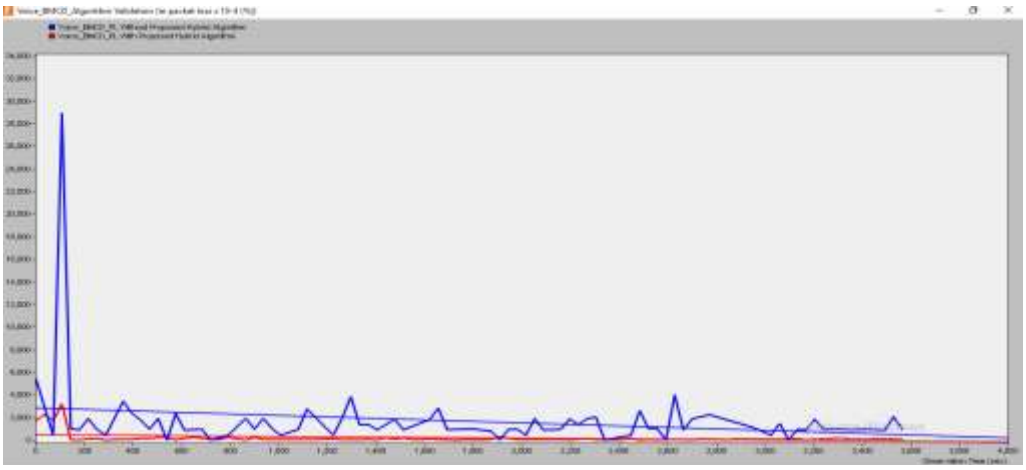


Figure 6: Network Packet Loss Probability with and without enabling Proposed Algorithm

Figure 7 shows the validation plot of packet losses against source intensity/offered load with and without of proposed hybrid algorithm with G.729 CODEC scheme. Packet loss probability fluctuates above zero percent with time when the QoS model is disabled but remains at zero percent when the model is enabled. This means that traffic provisioning will be highly unstable when the constrained-bandwidth network has default or zero scheduling algorithms. The proposed model therefore guarantees optimized QoS for voice and B/MCD traffics in terms of packet loss.

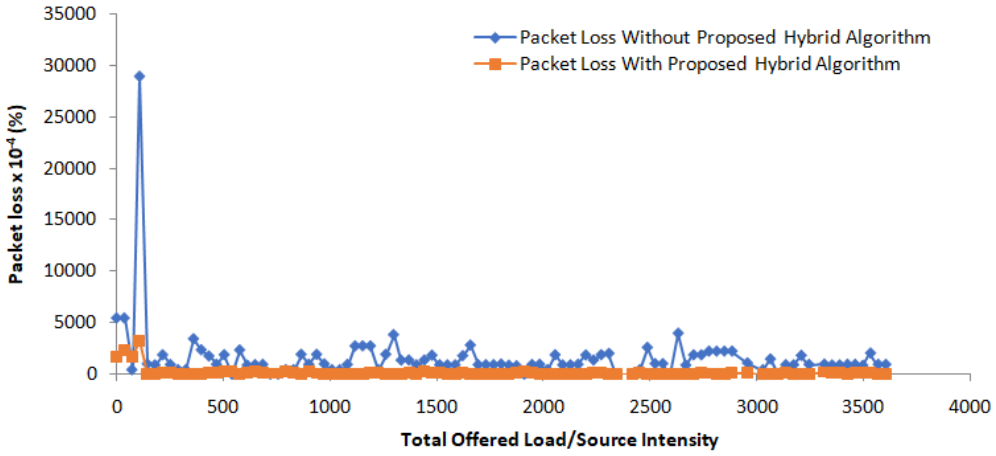


Figure 7: Validation plot of Packet Loss against Source Intensity/Offered load with and without the Proposed Algorithm Enabled.

VI. CONCLUSION

An optimized Voice and Business/Mission-Critical Data fair and priority queue scheduler for constrained-bandwidth VoIP networks has been designed, analyzed and simulated. Implementing the proposed model therefore ensures that every packet arriving the network is classified and explicitly marked for proper identification. Adequate precedence is given to both voice and business/mission critical data traffics. The provision of a dedicated (reserved) link handles the demands of the expected rapid increase in voice traffic. Excess voice and business/mission critical data packets that should have been lost are recovered and serviced with due priority. Other traffics (video and best-effort data) in the network are given fair treatment in the allocation of available resources. Fairness in resource sharing ensures that queues do not grow excessively, thereby reducing the delay and packet loss impairments. The optimized performance of the proposed scheme therefore guarantees a graceful tradeoff between priority (to voice and B/MCD traffics) and fairness (to all network traffics) in constrained-bandwidth VoIP networks without over provisioning the users. The design has been structured to be simple and easy to understand. It is developed in a modular form for easy manipulation, thereby making the scheduler architecture robust and consistent in its operation.

ACKNOWLEDGEMENT

This work is supported by the MacArthur Foundation Sub Grant, University of Port Harcourt.

REFERENCES

[1]. Cisco Systems, Understanding Delay in Packet Voice Networks, White Papers, Document ID: 5125, Cisco Systems Inc., San Jose, USA, 2006.
[2]. Y. Bandung, C. Machbub, A. Z. R. Langi and S. H. Supangkat, Optimizing Voice over Internet Protocol (VoIP) Networks based on Extended E-Model, Proc. IEEE Conference on Cybernetics and Intelligent Systems, Chengdu, 2008, 801-805.
[3]. J. M. Pitts and J. A. Schormans, Configuring IP QoS Mechanisms for Graceful Degradation of Real-Time Services, Proc. IEEE Military Communications Conference (MILCOM '06), Washington, DC, USA, 1-7, 2006, doi: 10.1109/MILCOM.2006.302111.
[4]. R. Shikhaliyev, Methods for Monitoring and Managing Computer Networks QoS, Proc. IEEE IV International Conference on Problems of Cybernetics and Informatics (PCI), Baku, Azerbaijan, 2012, 1-5.
[5]. L. Sun and E. C. Ifeachor, New models for perceived voice quality prediction and their Applications in Playout Buffer Optimization for VoIP Networks, Proc. IEEE International Conference on Communications, Paris, France, 2004, doi: 10.1109/ICC.2004.1312757.
[6]. L. Sun and E. C. Ifeachor, Voice Quality Prediction Models and their Applications in VoIP Networks, IEEE Transactions on Multimedia, Vol. 8, No. 4, 2006, 809-820.
[7]. C. Tu, Study on QoS Protection Mechanism of VoIP Systems, Proc. IEEE International Symposium on Intelligence Information Processing and Trusted Computing, 2011, 151-153.
[8]. Riverbed Technology, Riverbed Modeler: A Suite of Protocols and Technologies with a Sophisticated Development Environment, Riverbed Technology, San Francisco, USA. Available at: https://www.riverbed.com/gb/products/steelcentral/steelcentral-riverbed-modeler.html (Document accessed: Sept 5, 2018).
[9]. J. N. Dike and C. I. Ani, Algorithmic Analysis of an Efficient Packet Scheduler for Optimizing the QoS of VoIP Networks, International Journal of Computer Applications, Volume 97, Number 4, 2014, 5-12, doi: 10.5120/16993-7104.
[10]. P. Almquist, Type of service in the internet protocol suite, RFC (Proposed Standard) 1349, Internet Engineering Task Force, 1992.
[11]. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, An Architecture for Differentiated Services, RFC 2475, Internet Engineering Task Force, 1998:
[12]. D. D. Clark and J. Wroclaski,. An approach to service allocation in the Internet, Internet Engineering Task Force Internet Draft, 1997. Available at: http://tools.ietf.org/pdf/draft-clark-diff-svc-alloc-00.pdf (Document accessed: Sept 5, 2018).
[13]. K. Nichols, S. Blake, F. Baker and D. Black, Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers, RFC (Proposed Standard) 2474, Internet Engineering Task Force, 1998.
[14]. K. Nichols, V. Jacobson and L. Zhang, A two-bit Differentiated Services Architecture for the Internet, Network Working Group RFC 2638, Internet Engineering Task Force, 1999.
[15]. Cisco Systems, Implementing Quality of Service Policies with DSCP, Document ID: 10103, Cisco Systems Inc., San Jose, USA, 2008.
[16]. Cisco Systems, DIFFSERV-The Scalable End-to-End Quality of Service Model, Cisco White Paper 09186a00800a3e2f, Cisco Systems Inc., San Jose, USA, 2005.
[17]. R. Braden, L. Zhang, S. Berson, S. Herzog and S. Jamin, Resource Reservation Protocol (RSVP), Network Working Group RFC 2205, Internet Engineering Task Force, 1997.
[18]. Cisco Systems, Resource Reservation Protocol (RSVP), Cisco Systems Inc., San Jose, USA, Available at: https://www.cisco.com/c/en/us/products/ios-nx-os-software/resource-reservation-protocol-rsvp/index.html (Document accessed: Sept 5, 2018).
[19]. E. Rosen, A. Viswanathan and R. Callon, Multiprotocol Label Switching Architecture, Network Working Group RFC 3031, Internet Engineering Task Force, 2001.
[20]. A. Ma, Voice over IP (VoIP), SmartBits Performance Analysis System, Spirent Communications Inc., P/N 340-1158-001 REV A, 8/01, 2001. Available at: http://www.phonet.cz/archiv/dok_cizi/Spirent_100.pdf (Document accessed: Sept 5, 2018).
[21]. M. T. Gardner, V. S. Frost and D. W. Petr, Using Optimization to Achieve Efficient Quality of Service in VoIP Networks, Proc. IEEE International Conference on Performance, Computing and Communications (PCCC '03), 2003, 475-480, **doi:** 10.1109/PCCC.2003.1203732.
[22]. M. J. Fischer, D. M. B. Masi and J. F. Shortle, Approximating Low Latency Queuing Buffer Latency, Proc. IEEE Fourth Advanced International Conference on Telecommunications (AICT '08), 2008, 188-194, **doi:** 10.1109/AICT.2008.7.
[23]. M. E. Culverhouse, B. V. Ghita, P. Reynolds and X. Wang, Optimizing Quality of Service through the Controlled Aggregation of Traffic, Proc. IEEE International Conference for Internet Technology and Secured Transactions (ICITST '10), London, 2010, 1-7.
[24]. 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, Proc. IEEE National Conference on Communications (NCC '12), 2012, **doi:** 10.1109/NCC.2012.6176841.
[25]. K. Nisar, A. Amphawan, S. Hassan and N. I. Sarkar, A comprehensive survey on scheduler for VoIP over WLAN, Journal of Network and Computer Applications, vol. 36, no. 2, 2013, 933-948.
[26]. S. Ahmed, X. Jiang and S. Horiguchi, Efficient Scheduler for the Growing VoIP Traffic, Proc. IEEE International Conference on Parallel Processing Workshops (ICPPW '07), 2007, doi: 10.1109/ICPPW.2007.38.
[27]. K. Nisar, A. M. Said and H. Hasbullah, A Voice Priority Queue (VPQ) Fair Scheduler for the VoIP over WLANs, International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 2, 2011, 506-518.
[28]. A. Demers, S. Keshav and S. Shenker, Analysis and Simulation of a Fair Queueing Algorithm, Proc. ACM SIGCOMM Computer Communication Review, 19(4), 1989, 1-12, doi: 10.1145/75246.75248.
[29]. S. J. Golestani, A Self-Clocked Fair Queueing Scheme for Broadband Applications, Proc. IEEE Conference on Computer Communications, (INFOCOMM '94), 1994, 636-646, doi: 10.1109/INFCOM.1994.337677.
[30]. K. M. Chandy and C. H. Sauer, Approximate Methods for Analyzing Queueing Network Models of Computing Systems, Journal of ACM Computing Surveys, Vol. 10, No. 3, 1978, 281-317.
[31]. K. Salah, On the Deployment of VoIP in Ethernet Networks: Methodology and Case Study, Computer Communications, Vol. 29, Iss. 8, 2006, 1039-1054, doi: 10.1016/j.comcom.2005.06.004.

[32]. A. Leon-Garcia and I. Widjaja, Communication Networks: Fundamental Concepts and Key Architectures (McGraw-Hill Companies, Inc. New York, 2000).
[33]. J. D. C. Little and S. C. Graves, Little's Law in 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.
[34]. V. Jacobson, K. Nichols and K. Poduri, Expedited Forwarding PHB Group, Internet Engineering Task Force, 1999.
[35]. J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, Assured Forwarding PHB Group, Internet Engineering Task Force, 1999.
[36]. J. K. Sharma, Operations Research: Theory and Applications (Macmillan Publishers India Ltd., 5th Ed, 2013).

J. N. Dike "Design And Analysis Of Voice And Critical Data Priority Queue (VCDPQ) Scheduler For Constrained-Bandwidth Voip Networks "American Journal of Engineering Research (AJER), vol. 7, no. 10, 2018, pp. 154-167