International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.4 (July-2018)

http://dx.doi.org/10.12785/ijcds/070401

Adaptive Multi-Connection DASH Scalable Video Coding for Wireless Area Networks

Samar Ali[1], Yasser Ismail[2]and Ashraf Badawi[1]

[1]*The Center for Nanotechnology, Zewail City of Science and Technology, Cairo, Egypt*
[2]*Electrical Engineering Department, Southern University and A&M College, Baton Rouge, USA*

Received 2 Feb. 2018, Revised 7 May 2018, Accepted 14 Jun. 2018, Published 1 July 2018

Abstract: Dynamic Adaptive Streaming over Hypertext Transfer Protocol HTTP (DASH) has gained a significant momentum for multimedia streaming due to its ability in crossing firewalls and availability of infrastructure. In the mean time, Scalable Video Coding (SVC) is building a similar momentum as it enables efficient media storage and caching. In this work, we identify the main components of adaptive SVC-DASH client and propose a streaming heuristic over dynamic multiple connections. The proposed algorithm is experimentally tested under different connection and link configurations. Our results show that the algorithm successfully achieves interruption free streaming under all the tested BandWidth and link configurations. Additionally, the usage of multiple connections results in noticeable improvements in the achieved streaming quality for large link delays.

Keywords: Media Streaming, Video Encoding, Multiple Connections Enhancement, Layer Selection Policy

1. INTRODUCTION

Dynamic Adaptive Streaming over Hypertext Transfer Protocol HTTP (DASH) has gained a significant momentum and is widely adopted by many standardization bodies such as open Internet Protocol Television(IPTV) and the 3rd Generation Partnership Project (3GPP)[1, 2]. HTTP streaming has several attractive features that helped in speeding up its acceptance over other traditional streaming techniques such as Real-time Transport Protocol/User Datagram Protocol(RTP/UDP) [3, 4]. Avoiding firewall and Network Address Translation (NAT) traversal issues represent major advantages of HTTP streaming. Another advantage of using HTTP is the availability of HTTP cache infrastructures on the Internet that relieves not only the server load but also reduces the overall uplink traffic towards the cache[5, 6]. High end-to-end delay on the communication link (resulting from the limitation of the available transmission BandWidth (BW) is the main drawback of using HTTP streaming[7-9]. Adaptive streaming techniques are commonly used to avoid the limitation of the transmission BW. They accommodate the variations in the transmission BW in order to avoid annoying streaming interruption[10, 11]. Another problem of using HTTP streaming is the possibility of reducing the link efficiency due to Transmission Control Protocol (TCP) dynamics, however, this can be compensated by using a client-buffer[12].

Adaptive streaming techniques are commonly used to accommodate variations in the BW of transmission medium. As a result, annoying streaming interruption, due to buffer under-run, can be avoided. If Advanced Video Coding (AVC) based on H.264/MPEG-4 Part 10 standard is used, the server would contain multiple versions of each encoded segment at different qualities (resolution and frame rates) [6, 13]. Dynamic Adaptive Streaming over Scalable Video Coding (SVC-DASH) represents an extension to the AVC-DASH standard [10, 14]. SVC-DASH encodes the video segment into N layers including a base layer and numerous enhancement layers. The accumulation of more layers generally leads to an improvement of the video quality. This is because the granularity of quality control extends both horizontally over time and vertically over layers in SVC-DASH [15].

Many studies have focused on the design and the performance evaluation of video streaming over HTTP and using AVC - streaming over a single persistent connection. These studies considered different configurations of video encoding such as AVC and Scalable Video Coding (SVC) standards [10, 16, 17]. Others considered different connection configurations such as persistent versus non-persistent [18], single versus multiple connections, and streaming mode (real-time versus stored) and various segment durations [19].

SVC based streaming over HTTP has several merits that help avoiding the limitation of the BW of the

E-mail:samarali.cairouniv@gmail.com, yasser_ismail@subr.edu, abadawi@zewailcity.edu.eg

transmission medium as well as increasing the transmitted video quality. In[5], it is shown that SVC-DASH has several advantages including the possibility of serving a larger number of users with different equipment capabilities and a higher caching efficiency. In[12], SVC-DASH not only needs less buffering requirements in comparison to AVC-DASH but it also improves the Quality of Experience (QoE) for the viewer. In[10], the authors compare the performance of different AVC-based and SVC-based heuristics using NS3 simulations with NS3 cradle. The authors show that AVC performs better under high latencies while SVC better adapts to sudden and temporary BW fluctuations. In[20], the authors analytically investigate the optimal selection policy for layer segment when SVC is used. Their simulations show that a vertical policy is optimal under fixed BW while a diagonal selection policy is optimal for variable rate. In[19], the authors considers a cursor based SVC client heuristic Live TV setting where client side buffers is limited. The authors also exploit parallel and pipelined downloading of segment layers to overcome the high end-to-end delay issues. In another study[21], the authors present a rate adaptive algorithm for AVC-based video streaming over two parallel connections. Their simulations show that parallel connections outperforms the serial segment fetching method in achievable media bit-rates but slightly inferior in buffer underflow frequency. In most of the previously discussed work, high speed video streaming is the most important target. However, getting high speed video streaming is still an issue.

In this paper, we are targeting a high speed video streaming transmission. A novel SVC-DASH client over Multiple parallel connections (SVC-DASH-M) system is implemented and evaluated. In SVC-DASH-M, the quality of the requested layer-segment depends on the amount of buffered media and the network condition. Additionally, it adopts two selection policies that are conservative for the frequency of the video quality shifts.

The paper is organized as follows, in section 2, SVC-DASH-M system is discussed in details. Implementation and performance evaluation is discussed in section 3. A conclusion will be drawn in section 4.

2. SVC-DASH-M SYSTEM

The top level structure of the proposed SVC-DASH-M system is shown in Figure 1.The raw video data is subdivided into short duration chunks, commonly known as segments by the JSVM encoder according to the requirements of the HTTP streaming[22]. Every video segment is split into multiple layers per segment depending on the required video quality level, image size, and frame rate. The contents of each layer segment are stored as a separate file on the HTTP server. The details of the video information are typically stored in a media presentation description (MPD) file that includes web links to the media files with the corresponding encoding details of each file contents [1].The MPD file contains some information such as segment ID, layer ID, layer-segment URL, layer-segment duration, and layer size. It is worth mentioning that the segment ID can be used to identify the timing information of the transmitted segment. In HTTP streaming, the content is requested using HTTP GET request/response dialog that is typically serviced over TCP. Once the DASH client requests a specific video, the streaming process starts by downloading the MPD file through the HTTP streaming client. The MPD parser is parsing the MPD file to obtain the video information.

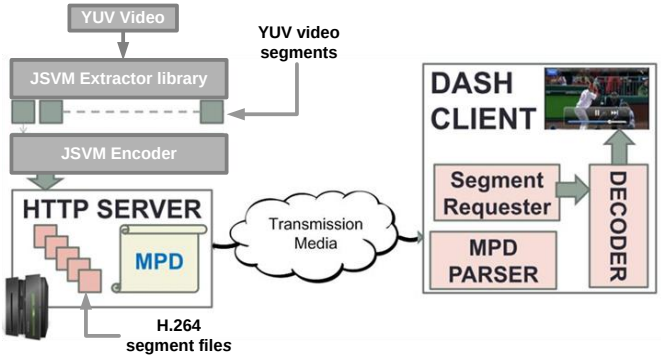


Figure 1. SVC-DASH-M System.

Once the MPD information is received, our multiple-connection client requests the first n_{min} base layer-segments via the segment requester, where n_{min} represents the minimum number of connections. Upon the reception of a complete layer-segment on connection c, SVC-DASH-M considers a level-based layer-segment requester that determines the quality of the requested layer-segment.After the reception of a segment with the available layers, it can be decoded and viewed. The proposed DASH client algorithm and its advantages are summarized in the following paragraph.

DASH client algorithm defines two application buffer-level thresholds denoted as B_{min} and B_{target} . B_{min} represents a low threshold for the data to be maintained in the buffer to avoid streaming interruptions while B_{target} represents a target buffer level that the application should be operating around. Improving the streaming quality at high buffer levels through downloading enhancement (successive) layers is the main advantage of our DASH client algorithm. Additionally, in case of streaming interruptions, our algorithm maintains the user experience quality at low buffer-level by focusing on the base layer-segments. The dynamics of the used connections are considered as an additional critical design advantage of the proposed SVC-DASH-M. In our

algorithm, the number of opened connections, denoted as n_c , is lower-bounded by n_c^{min} and upper-bounded by n_c^{max}. The number of used connections varies depending on the streaming performance of connection c over which the segment is received. In the following subsection, the detailed components of SVC-DASH-M system are discussed in more details.

A. HTTP DASH Server

HTTP DASH server contains the MPD file and the layer segment files for each video. The layer segment files are generated as follows (Please see Figure 1):

1.The original YUV video file is sliced into smaller YUV video segments using The JSVM BitStream Extractor library [22].

2.Each small video segment is encoded using JSVM encoder, which generates the encoded H.264 segment files for each video.

3.Each H.264 segment file is processed to extract the bytes corresponding to each layer as shown in Figure 2. This extraction is performed using a C code; developed for this purpose as none of the existing tools provide this function.

The layer segment frames extraction process, shown in Figure 2, is summarized as follows. Given the H.264 segment file (Seg_file), and the Layer_info.txt layer file, the start position, the end position, and the size of each frame constituting a layer of a certain segment is fetched from the Layer_info.txt layer file. The layer frame bytes will be then fetched from the segment file (seg_file). Number of bytes should equal to the frame size. The data will be then written into the H.264 layer file. The process will be repeated until the bytes of each frame constructing certain layer are moved to the H.264 layer segment file. This process is performed for each frame constituting a segment layer.

The MPD file is prepared, as seen in Figure 3, as follows. The JSVM BitStream Extractor compiles the H.264 segment file and produces the segment trace file. The segment trace file has the start position, size, scalability levels of each Network Abstraction Layer Unit(NALU), the packet type (stream header or slice data), and the packet importance (Discardable or Truncatable). The frame is consisted from multiple NALU.Using the AWK programming language, the frames constituting complete layer is extracted and their related information is written in a separate .txt file. The bit-rate of each layer is also extracted using the JSVM BitStream Extractor that calculates the accumulative bit-rate of each video quality layer. The segment ID, Layer ID, Layer size in bytes, segment duration in seconds, the URL of each layer, and the layer bit-rate is written to the MPD file to provide the streaming client with important information required for each segment layer and its associated file. All the segment layer files together with the MPD file are stored on the HTTP server.

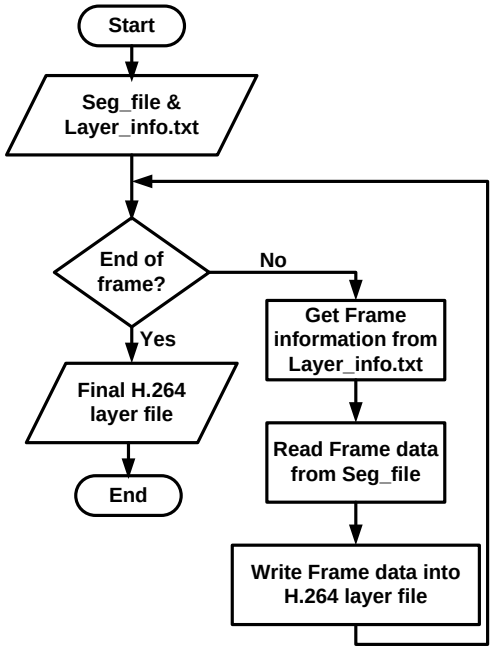


Figure 2. Extraction of H.264 Layer file.

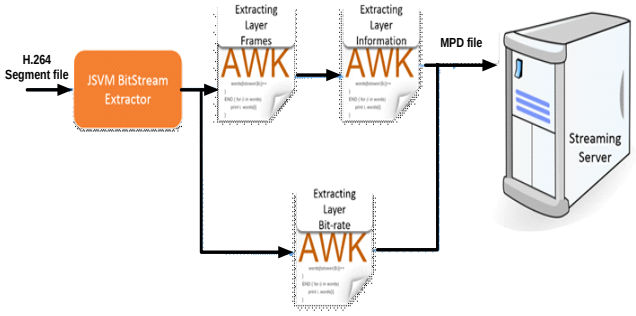


Figure 3. MPD preparation process.

B. Dash Client

Typically, the first step performed by an HTTP streaming client is to download the MPD file and parse it to obtain the video information. Once the MPD information is received, SVC-DASH-M initiates the operation based on three interacting rules. First, the Layer-Segment Requester that determines whether a base layer or an enhancement layer to be downloaded. The Connection Manager Design and the Enhancement Layer Selection policy are the remaining two rules. Table I illustrates the main functions implemented in the client algorithm. Followings are some detailed description of each function.

TABLE I. CLIENT ALGORITHM FUNCTIONS.

Function	Meaning
GetBuffLevel()	Counts the number of segments in the buffer, and returns the buffer level in seconds.
reqLaySeg(lsType, conn)	Requests a layer-segment from the server over connection conn. IsType could be "B" for base layer, "E" for enhancement layer, or "A" to alternate between base and enhancement layers.
estimateNextRate(lsType)	Takes the layer segment type, and returns the data rate required for downloading the next two segments to be requested.
reqNextLaySeg(lsType, conn)	Given the layer-segment type, this function requests the next layer-segment to be requested over conn. If conn is set to -1, then a new connection is opened to request the selected layer-segment. Or the conn is set to the number of already opened connection to request a layer segment.

i. Layer-Segment Requester

The layer-segment requester initially gets the first n_{min} base layer-segments, where n_c^{min} represents the minimum number of connections. Upon the reception of a complete layer-segment on connection c, SVC-DASH-M considers a level-based layer-segment requester that determines the quality of the requested layer-segment as shown in Algorithm1.

Algorithm 1: Level-based Layer-Segment Requester Algorithm

bl = getBuffLevel(); //buffer level
if bl<= B_{min}
//request base layer only
reqLaySeg(B,conn);
elseif B_{min}<bl<= B_{target}
//alternate base and enhancement requests
reqLaySeg(A,conn);
else
//request enhancement layers only
reqLaySeg(E,conn);
end if

The details of getBuffL(), and reqLaySeg() functions are shown in table I. The algorithm defines two application buffer-level thresholds denoted as B_{min} and B_{target}. B_{min} represents a low threshold for the data to be maintained in the buffer to accommodate network condition variations. B_{target} represents a target buffer level that the application should be operating around. The defined policy in Algorithm 1 targeting improving the streaming quality at high buffer levels through downloading enhancement layers. Additionally, the layer-segment requester policy maintains the user experience quality at low buffer-level through avoiding streaming interruptions by focusing on the base layer segments.

ii. Connection Manager Design

The connection manager controls the dynamics of the used connections in SVC-DASH-M. In the implemented algorithm, the number of opened connections, denoted as n_c, is lower bounded by n_c^{min} and upper-bounded by n_c^{max}. The number of the used connections varies when a layer-segment is received. Let c denotes the connection over which the layer-segment is received. The connection manager operation is shown in Algorithm 2. The details of the used functions are presented in table I. The streaming performance is evaluated based on the segment duration-fetch ratio $\mu = \frac{SD}{SFT}$, where SD represents the duration of the received segment and SFT represents the duration over which the segment is fetched. Note that large values of μ indicate that the available BW is large and vice verse. This is because a new layer-segment is requested over. An additional segment may be requested over a new connection if $n_c < n_c^{max}$ and there exists sufficient BW for downloading two segments. The BW sufficiency condition is satisfied if the ratio between the required BW for downloading the two selected layer-segments (denoted as r_{nxt}) to the estimated connection BW (denoted as r_{prev}) is less than the segment duration-fetch ratio. To this end, it is worth noting that all the decision parameters are available or are readily measured at the application layer and do not incur any additional overhead to the communication process.

Algorithm 2: Connection Management Algorithm

$\mu = \frac{SD}{SFT}$;
r_{nxt} = estimateNextRate(lsType);
ε =(r_{nxt} / r_{prev}) ;
if μ >1+ ε
reqNextLaySeg(lsType, conn);
if (n_c<n_c^{max})
reqNextLaySeg(lsType, -1);
end if
elseif μ <1
ifn_c>n_c^{min}
close(conn);
else
reqNextLaySeg(lsType, conn);
end if
else
reqNextLaySeg(lsType, conn);
end if

C. Enhancement Layer Selection policy

The granularity of the quality control extends both horizontally over time and vertically over layers. The enhancement layer selection policy improves the current segment quality in conjunction with keeping consistent quality level for future segments, so that a balance between the current segment quality and the quality of the future segments has to be guaranteed. It should be ensured that any selected layer-segment is expected to be received before its play out time. Additionally, the requested quality level of the segment should be brought to the buffer before the play out time of this segment, otherwise, the download of the enhancement layer of this segment will be no use and the segment is displayed with the enhancement layers that already existed in the buffer or as a base layer if none of its quality levels is downloaded.

i. Vertical Policy

The vertical enhancement layer selection policy concerns the quality of the current segment and doesn't consider the next segment quality. The streaming client downloads all the enhancement layers of the current segment then the next segment and so on. As shown from Figure 4, for the vertical layer selection policy, all enhancement layer-segments are downloaded for segment i before downloading enhancement layer-segments for segment i+1 , which is greedy in nature from the future segments perspective. As the vertical policy cares about the current segment quality, in order to keep the same quality for the future segments buffer starvation may occur. If the client wants to avoid the buffer under-run, quality variation is happened that may be annoyed.

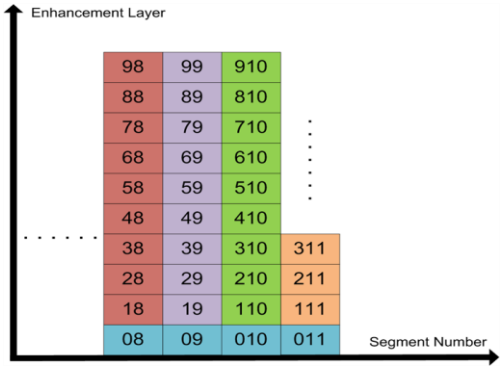


Figure 4. Vertical Layer Selection Policy.

ii. Horizontal Policy

The horizontal layer selection policy minimizes the quality variations among the segments, the streaming client downloads certain enhancement layer for all the segments already in the buffer, then move to downloading higher quality level and so on. As shown from Figure 5 that the horizontal layer selection policy ensures that no layer-segment from layer n is requested unless all the layer-segments of layer n-1 have been already downloaded. This policy prevents requesting higher quality-layer segments if there is not enough BW, it aims to reduce the quality shifts in the received video.

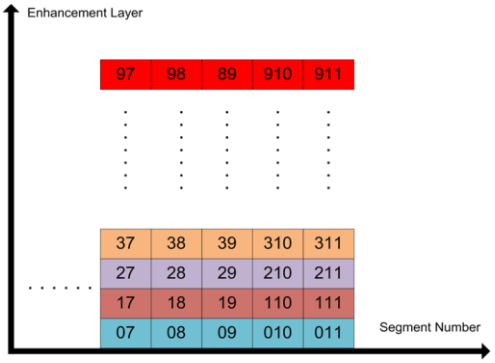


Figure 5. Horizontal Layer Selection Policy.

D. Streaming Algorithm Implementation

The proposed streaming client algorithm starts by receiving a URL for the Media Presentation Description file (MPD). The client then initializes the streaming process by:

a) Requesting the MPD from the HTTP server.

b) Parsing the received file to extract the information about the video including segment Id, layer Id, size of layer in bytes, segment duration, available quality per segment, URL, and layer bitrate in Kb/s.

c) Storing these information in an array of structures, each element of that array has these information in addition to the playout time of each segment in the video.

d) Request a number of segments equals to minimum number of segments or the minimum number of the connections.

The client then starts a loop in which it downloads segments and delivers them to the application buffer until the stopping is satisfied. The stopping condition in our case is the playout of the last segment after which the download of any additional segment layer would be of no benefit. In this segment, the client continuously receive from the open sockets and update the buffer level. the next step is to update the buffer level by the difference duration of the delivered segments and the played duration. The played duration is estimated as the difference between the current time and the sum of both the play out start time and the accumulated interruption delay due to buffer under-run if any.

Executing the HTTP streaming algorithm. The algorithm defines different operating modes depending on the current buffer level and the network condition. The network condition depends on two parameters. The first parameter is the network indicator ratio $\mu = \frac{SD}{SFT}$, where SD represents the duration of the received segment and SFT represents the duration over which that segment is fetched. Large values of μ indicate that the available BW

is large and vice versa. The second parameter is the application demand ratio $\&e_{psi} = r_{next} - r_{prev} \times r_{cur}$, where r_{next} represents the rate of the next two segments to be requested and r_{prev} represents rate of the received segment. Large values indicate that changing requesting additional segment would require large additional BW. The network is considered in a good condition if $\mu > 1 + \&e_{psi}$, and the network is considered in a bad condition if $\mu < 1$.

Two buffer levels are defined including MIN, and TARGET levels. The MIN level represents the initial playout latency that is typically maintained in streaming applications to accommodate the impact of variations in link conditions. The TARGET buffer level represents a target buffer level that the application should be operating around in order to focus on improving the quality of the video. The choice of these levels would be investigated during the testing phase. Upon the reception of a complete segment, the client estimates the network condition and checks the buffer level and enters one of the operating modes accordingly as described below:

i. Upon the reception of a segment, the application do the following :
   - If the segment is out of order, it is delivered to an ordered segment waiting list.
   - If the segment is received in order, it is delivered to the application buffer and the buffer level is updated. Additionally, the waiting segments are checked for other possible deliveries.

ii. Then, the network indicator ratio is calculated.
   - If the buffer level is below MIN, the application demand factor is estimated for the next two base layer segments.
   - If the network is good, an additional connection is opened and one base layer is requested on each of the available two connections (the one that is just opened and the one on which the last segment is received). If a new connection cannot be opened (the current number of connections equals the per-defined maximum number of connections), a base layer is requested on the current connection.
   - If the network is bad, close the current connection if the opened connections is more than the minimum number of connections. otherwise request the next base layer segment on this connection.
   - If the network conditions is neither good nor bad, a new base layer segment is requested on an empty connection.

iii. If the buffer level is above the MIN level and below the TARGET level, the application demand factor is estimated for the next base layer segment and the next enhancement segment. The order of the enhancement segments is determined based on a per-defined policy. For example, the policy may be a horizontal policy such that segments belonging to layer n will not be requested until all lower layer segments are downloaded.
   - If the network is good, an additional connection is opened and the next base and enhancement layers are requested on the available connections. If a new connection cannot be opened, a base layer segment is requested if the number of active connection downloading a base layer segment is less than or equal the maximum number of connections. Otherwise, an enhancement layer segment is requested.
   - If the network is bad, close the current connection if the opened connections is more than the minimum number of connections. otherwise request the next base layer segment on this connection.
   - If the network is neither good nor bad, a new base layer segment is requested on an empty connection.

iv. If the buffer level is above the TARGET level, the application demand factor is estimated for the next two enhancement segments.
   - If the network is good, an additional connection is opened and the next base and enhancement layers are requested on the available connections. (if there are no base layer for any segment to be downloaded so the two enhancement layers will be requested).
   - If the network is bad, close the current connection if the opened connections are more than the minimum number of connections. otherwise request the next enhancement layer segment on this connection.

If the network is neither good nor bad, a new enhancement layer segment is requested on an empty connection.

3. Implementation and Performance Evaluation

A. System Setup

A real testbed network composed of a server, a transmission media, and a client are implemented and emulated to test the proposed algorithm, as seen in Figure 6.

Figure 6. Testbed Architecture.

i. The streaming DASH Server

The DASH server is a typical laptop with Ubuntu 12.04 LTS and Apache[23]as a standard HTTP server. In our evaluation, the YUV videos used for testing are obtained from Arizona State University video library[24], table II shows the details of the used YUV videos in our testing.

TABLE II. THE USED YUV VIDEO SEQUENCES.

Video	# of frames	Frame Rate (fps)	Duration in seconds	# of segments
Paris	1065	24	44.375	23
Big Bunny	2880	24	120	60
Highway	2000	24	83.33	42

Each video is processed using the Joint Scalable Video Model (JSVM) [22]as following :

1. The original YUV video file is sliced into smaller YUVvideo segments using JSVM down-sampler. The duration of the segment is a design parameter that is selectedas 2 sec.
2. Each segment is encoded to ten scalable layers including spatial (original CIF and a down-sampled QCIF), temporal and quality scalability.
3. The encoder produces one H.264 file per segment. An additional code was developed to extract the different portions corresponding to different layers in each segment using the layer information obtained from JSVM bit-stream-extractor.

The corresponding MPD file is then compiled based on the extracted layer-segment information. Finally, the MPD file together with the layer segment files are uploaded to our HTTP DASH server.

ii. The emulated network

Traffic control tools are employed to emulate different BW and delay values. Netem is used to set the delay. Token bucket filter (tbf) is used to set the link BW to the values of interest. For the proper setup of tbf, we set: $\text{limit} \geq \text{BW} * \text{HZ}$ and $\text{buffer(burst)} = \frac{\text{BW}}{\text{HZ}}$, whereBWrepresents the emulated BW and HZrepresents theLinux kernel timer interrupt frequency. It is worth mentioning that one may need to reset scatter-gather, TCP-segmentation-offload, generic-segmentation-offload, and generic-receive-offload using ethtoolto avoid packet dropping.

iii. The client

The client is a laptop with linux OS (Ubuntu, 12.04 LTS). Our client is implemented using C programming language to connect and fetch the MPD file and video layer-segments according to the presented algorithm. The implemented C code takes the URL of the MPD file, the minimum, and the maximum number of connections (n_c^{min} , n_c^{max}) as inputs. Table III and table IV illustrate the parameters of our client buffer, where the minimum buffer level represents the play-out latency that is typically maintained to accommodate the impact of variations in link conditions. The application should be operating around the target buffer level to focus on improving the video quality. The application has two operating modes. The streaming client enters the first mode in two cases:

- At the start of the playout, the application has to download a number of segments equals the minimum buffer level.
- If the buffer level is depleted (reached zero seconds) during the streaming, the application re-enters a buffering mode until it secures a six second portion in the buffer again.

The second mode is defined as the streaming mode. During such mode, the client is streaming the downloaded media without any interruptions even if the buffer level is below the minimum. Wireshark is used to trace the transmitted packets to estimate some performance metrics. Then the received video layer segments are decoded and the PSNR is calculated. Note that the average quality is estimated over all received segments. These performance parameters are evaluated for different link and delay configurations.

TABLE III. THE SEGMENTS USED FOR THE BUFFER LEVEL.

The Buffer Level	Segments	Seconds
The Min. Buffer Level	3	6
The Target Buffer Level	5	10

TABLE IV. THE USED PARAMETERS OF THE USED CLIENT BUFFER.

Index	Definition
$\mathbf{n_i}$	The number of interrupts
$\mathbf{n_c}$	The number of opened connections
$\mathbf{n_{cc}}$	The number of closed connections due to the low BW
$\mathbf{d_v}$	The application downloaded data in Bytes.
$\mathbf{t_d}$	The time at which the application stops downloading more layer-segments
$\boldsymbol{\gamma}$	The application goodput that is estimated here as the ratio between the total transmitted video layer-segment sizes and the time spent in transmitting this amount of data.
q	The average quality in terms of the number of layers (i.e, As the number of layers increase, the average quality increases)
PSNR	Peak-signal-to-noise-Ratio, The PSNR is calculated on the frame level by comparing each pixel of every frame in the reference video and the received one.

B. Simulation Results

We investigate the impact of both the horizontal and diagonal enhancement layer selection policies on our SVC-DASH-M streaming client when our client using the whole BW, and under real environment when there is cross traffics with the DASH client. The evaluation of the received video is done based on the video quality and the PSNR. The video quality is measured as the number of

layers achieved. Experiments are done under several network conditions for different connection configurations when the video is streamed.

i. Emulation with no crosstraffic

Figure 7 plots the average received quality for the streamed Paris video versus different BW configurations for different min-max connection configurations at l0ms link delay. The results represent the average quality for all video segments over five runs. Intuitively, increasing the BW improves the quality of the received video for all connection configurations. Additionally, it is clear from Figure 7 that a single connection represents a good choice for the tested link configurations especially at low BW values. We also noticed that in case of using a large maximum number of connections, the algorithm closes many connections as their segment fetch time is larger than their segment duration especially at low BW. This observation implies that opening too many connections may not be useful as they incur additional overhead without other gains.

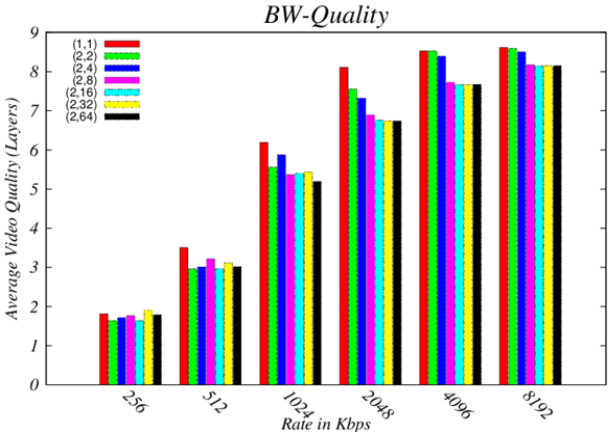


Figure 7. Average Video Quality Versus BandWidth for different connection configurations (n_c^{min} , n_c^{max}) for the Horizontal Policy.

Figure 8 shows the achieved quality for each segment considering different connection configurations for an average link delay of 10 ms and BW of 8 Mbps. The figure provides us with some insights on the quality dynamics versus time. The figure shows that all the connection configurations managed to download the highest quality towards the video session end. However, the main difference lies in the starting dynamics. At the beginning of the streaming, the adopted conservative approach for layer segment request implies requesting base layer segments only and limits the benefit from the available BW to download enhancement layer segments. Note that this conservative approach helps reducing the initial playout latency, which is another important QoS metric.

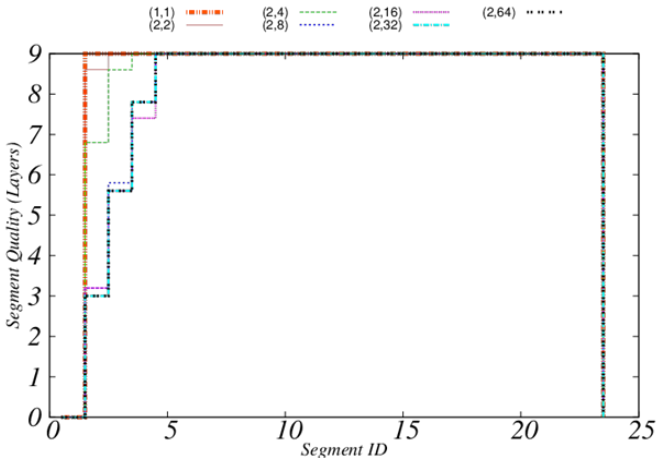


Figure 8. Segment Quality versus segment number for different connection configurations(n_c^{min} , n_c^{max}) for the Horizontal Policy.

Table V shows some of the several KPIs for different link BWs with a 10 ms link delay. As a general observation, the connection configuration has limited impact on all the estimated KPIs except for $n_c$and n_{cc}. For any configuration, the algorithm opens more connections as the upper-bound n^{max} increases. It is shown from Table V(a) that approximately 12 connections out of 19 opened connection for (2,32) configuration are closed. It is also shown that increasing the value of n^{max}above a certain limit does not have any effect on improving the performance, and this limit depends on the link parameters.

TABLE V. KPIS FOR DIFFERENT LINK BANDWIDTHS.

(a) 256Kbps BandWidth							
(n_c^{min} , n_c^{max})	**(1,1)**	**(2,2)**	**(2,4)**	**(2,8)**	**(2,16)**	**(2,32)**	**(2,64)**
n_i	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n_c	1.00	3.00	13.80	18.60	17.20	19.40	18.40
n_{cc}	0.00	0.00	7.40	11.40	10.80	12.20	11.80
d_v(MB)	1.5	1.3	1.4	1.4	1.3	1.5	1.4
t_d(sec)	37.74	36.81	38.89	41.12	37.89	40.69	40.34
y(Kbps)	39.14	35.82	35.74	34.24	35.35	36.35	34.85
q	1.81	1.63	1.71	1.77	1.63	1.90	1.78
(b)1MbpsBandWidth							
(n_c^{min} , n_c^{max})	**(1,1)**	**(2,2)**	**(2,4)**	**(2,8)**	**(2,16)**	**(2,32)**	**(2,64)**
n_i	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n_c	2.00	3.20	14.40	24.80	40.60	45.60	42.00
n_{cc}	0.00	0.00	6.40	11.60	20.20	16.20	16.60
d_v(MB)	5.5	5.1	5.2	4.8	4.66	4.72	4.74
t_d(sec)	37.82	35.01	35.65	32.32	31.09	31.50	31.66
y(Kbps)	145.28	145.31	146.33	148.39	149.92	149.76	149.64
q	6.19	5.56	5.87	5.37	5.40	5.43	5.19
(c) 2Mbps BandWidth							
(n_c^{min} , n_c^{max})	**(1,1)**	**(2,2)**	**(2,4)**	**(2,8)**	**(2,16)**	**(2,32)**	**(2,64)**
n_i	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n_c	4.60	3.40	10.80	29.60	50.00	57.40	56.20
n_{cc}	0.00	0.00	4.80	9.80	14.20	16.80	18.80
d_v(MB)	6.99	6.58	6.45	6.03	5.85	5.83	5.73

t_d(sec)	33.17	23.44	22.89	21.23	20.50	20.38	20.01
y(Kbps)	228.44	280.85	281.75	284.20	285.39	285.88	289.63
q	8.03	7.42	7.12	6.85	6.79	6.77	6.77
(d) 4Mbps BandWidth							
(n_c^{min}, n_c^{max})	**(1,1)**	**(2,2)**	**(2,4)**	**(2,8)**	**(2,16)**	**(2,32)**	**(2,64)**
n_i	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n_c	3.00	4.00	5.40	19.00	60.00	66.80	67.00
n_{cc}	0.00	0.00	0.40	4.80	18.20	19.00	21.20
d_v(MB)	7.11	7.11	6.98	6.56	6.41	6.41	6.39
t_d(sec)	12.86	12.84	12.58	11.73	11.43	11.44	11.40
y(Kbps)	533.08	533.76	555.15	559.47	560.71	560.44	560.38
q	8.52	8.52	8.39	7.72	7.66	7.66	7.67
(e) 8Mbps BandWidth							
(n_c^{min}, n_c^{max})	**(1,1)**	**(2,2)**	**(2,4)**	**(2,8)**	**(2,16)**	**(2,32)**	**(2,64)**
n_i	0.00	0.00	0.00	0.00	0.00	0.00	0.00
n_c	3.00	4.00	8.40	14.60	52.60	58.40	61.20
n_{cc}	0.00	0.00	0.40	1.20	1.00	2.80	1.40
d_v(MB)	7.12	7.12	7.08	6.77	6.77	6.72	6.72
t_d(sec)	6.53	6.54	6.45	6.10	6.17	6.12	6.12
y(Kbps)	1.1	1.1	1.1	1.1	1.1	1.1	1.1
q	8.61	8.59	8.50	8.17	8.14	8.15	8.15

Our results indicate that under all the tested configurations, the algorithm manages to adapt the streaming to different BWs and no interrupts are encountered. Also, as the link BW increases, the application throughput increases as shown in Figure 9. This increase is due to the ability to download more enhancement layer-segments and the typical drop of the download time (t_d). Consequently, a higher average quality is observed as the BW increases. It is also shown from Figure 9 that all the connection configurations achieve approximate throughput under the same network BW. It is worth mentioning that such download dynamics affects the decisions of closing and opening new connections and hence, using larger number of connections may limit the benefit of persistent connections. On the contrary, using a smaller number of connections reduces the need of simultaneous connections opening and also avoiding splitting of the BW, as a result, the segment download time is relatively small.

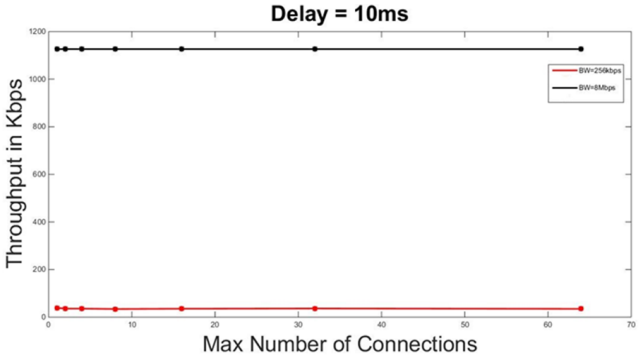


Figure 9. Throughput versus maximum number of connections for different networks.

ii. Emulation with crosstraffic

In order to emulate the real environment, some cross traffic is intentionally added as shown in Figure10. The iperf tool is used to make a Constant Bit Rate (CBR) UDP traffic while the streaming client is running. As the iperf client is the data source and the iperf server is the data sink, consequently, the DASH streaming client device is set to be iperf server also. Additionally, the DASH server is acting as the iperf client in order to have another traffic during the streaming route. Two scenarios are simulated. First, when the cross traffic is starting after the DASH streaming client by 6 seconds (i.e., the play out latency period), and lasts for 10 seconds. This experiment is tested when the cross traffic BW is equal to the settled BW as indicated from Figure 11.a, and when the cross traffic BW is greater than the established BW as shown in Figure 11.b. The second scenario, as shown in Figure12, is done when the cross traffic is existed during the streaming period. This experiment is done for the cross traffic BW equals 50% of the settled BW, and for the cross traffic BW equals 90% of the settled BW. Both the two scenarios are tested for the Paris video using the horizontal and diagonal enhancement layers election policies.

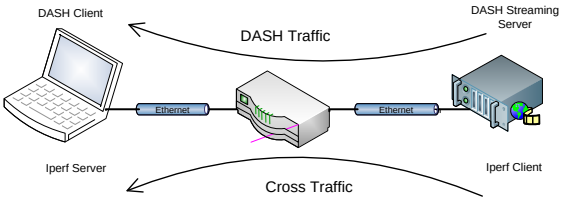
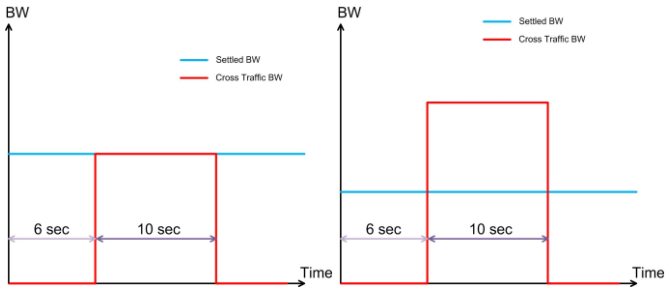
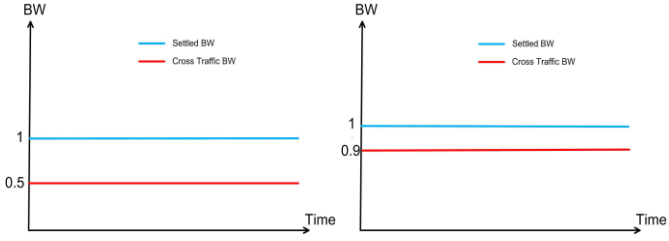


Figure 10. Setupforthecrosstrafficexperiment.



(a)Cross Traffic BW=Settled BW (b)Cross Traffic BW=2*Settled BW

Figure 11. SettledBWvs.CrossTrafficBW.



(a) Cross Traffic BW=50% of Settled BW (b) Cross Traffic BW=90% of Settled BW

Figure 12. Settled BW vs. Cross Traffic BW when The cross traffic is existed during the streaming period.

Figure 13 shows the achieved quality for each segment considering different connection configurations for an average link delay of 10ms, a BW of 2Mbps, and a CBR UDP cross traffic of 2Mbps using the horizontal layers election policy. It is shown that the streaming client implementing the algorithm is free of interruptions for all connection configurations. It is also indicated that as the maximum number of connections increase morethan2, the achievable quality for the first segments is low compared to the cases of the (1,1) and (2,2) connection configurations, this is because of the division of the BW among the opened connections. The segment quality suddenly drops to the first for the single connection, which means that the multiple connections are resistant to the cross traffic when comparing with the single connection case. After the end of the cross traffic, most of the connection configuration can achieve the maximum quality except the cases of (2,32), and (2,64), this is due to the fact that when the number of connections increases, each connection takes a small portion of the BW.

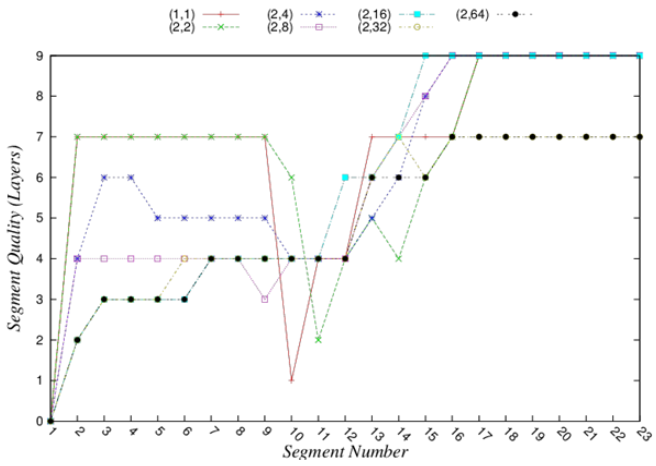


Figure 13. Segment Quality Vs. Segment Number at 2 Mbps UDP cross traffic for different connection configurations (n_c^{min} , n_c^{max}) for the Horizontal Policy.

Figure 14 illustrates the achieved quality for each segment considering different connection configurations for an average link delay of 10ms, a BW of 2Mbps, and a UDP traffic with rate 4Mbps. It is clear from Figure 14 that as the maximum number of connections increases, the interruptions disappeared. The algorithm is interruption free for the maximum number connections above2, and all the segments are displayed between quality levels 3 and 4 during the cross traffic period. This indicates that as the number of parallel working connections increase, this gives more immunity to the cross traffic. It is also demonstrated that the single connection records a 5 second interruption after the download of segment 11. This is because when there is only one opened connection and the network conditions deteriorate suddenly. Consequently, this breaking down has a great negative effect on the only one opened connection. And also for the (2,2) connection configurations, there is interruption lasts for 3 second after the download of the 12th segment. While for the multiple connection case, the network variations is divided among the opened connections. For illustration, figure shows the average video quality when the cross traffic BW starts after 6 sec from the streaming, one time at the cross traffic BW equals the settled BW, and the other at the cross traffic BW doubles the settled BW.

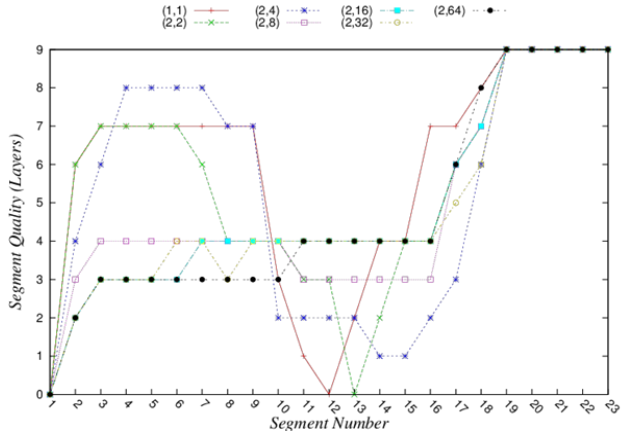


Figure 14. Segment Quality Vs. Segment Number at 4 Mbps CBR UDP cross traffic for different connection configurations (n_c^{min} , n_c^{max}) for the Horizontal Policy.

Figure 15 shows the average video quality at the cross traffic BW equals 50%, and 90% of the settled BW for an average link delay of 10ms, a BW of 2 Mbps for different connection configurations. Figure 15 shows that as the cross traffic BW increases, the average quality decreases. The algorithm is interruption free for the different tested connection configurations. It is noticed that the average quality increases as the maximum number of connections increases then saturates.

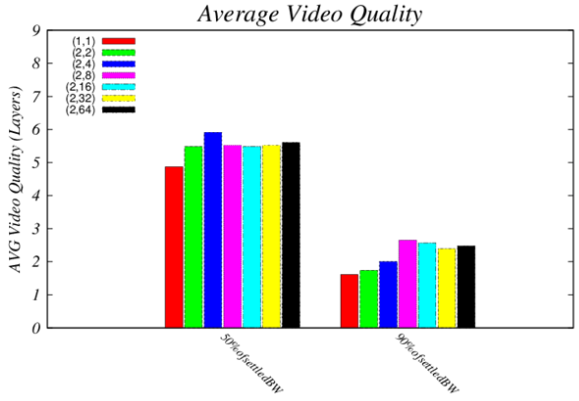


Figure 15. Average Video Quality at the cross traffic equals 50%, and 90% of the settled BW for different connection configurations (n_c^{min} , n_c^{max}) for the Horizontal Policy.

Figure 16 shows the achieved quality for each segment considering different connection configurations for an average link delay of 10ms, a BW of 2Mbps, and a CBR UDP cross traffic of 2 Mbps using the diagonal enhancement layer selection policy with slope=1. In comparison with the horizontal policy, the figure shows that the achieved segment quality is higher for all connection configurations, and the sudden quality shifts are reduced for the single connection case which reduce the user's disturbance as the quality of the video segments are smoothly changing. After the cross traffic removal after the 16th second from the start of streaming, all the connection configurations succeed in reaching the maximum quality.

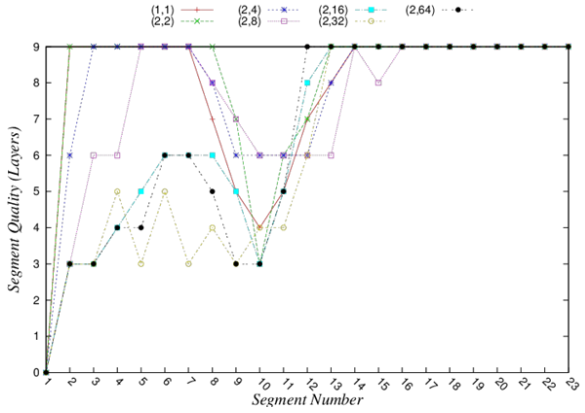


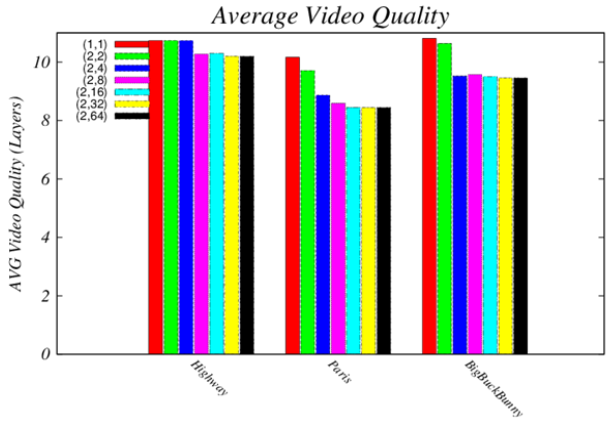
Figure 16. Segment Quality Vs. Segment Number at 2 Mbps CBR UDP cross traffic for different connection configurations (n_c^{min} , n_c^{max}) for the Diagonal Policy.

Figure 17 shows the achieved quality for each segment considering different connection configurations for an average link delay of 10ms, a BW of 2Mbps, and a CBR UDP cross traffic with rate4 Mbps. The results show that interruptions occurred for the single connection, (2,2), and (2,4) as the diagonal policy concerns improving the quality for the current and future segments. For example, for the (2,2) connection configuration, the download of the enhancement layers for segments 10, and 11 is the cause for the buffer under run for 5 seconds. When the maximum number of connections increases more than 4, no interruptions occurred.

Three different videos are encoded into 12 scalable layers from Base layer to layer 11, and tested using our DASH client, the average of each video sequence is calculated. Then, using Open SVC Decoder for decoding and JSVM PSNR library for PSNR calculation. Table VI provides some information about the tested video sequences. Figure 17 plots the average video quality for different video sequences using different connection configurations for a BW of 2 Mbps, and an average link delay of 10ms. The network BW is chosen to be more than the bitrate of any tested video. The average quality for high way is the highest as it has less information, also its bitrate is less than the BW. As a result, the video average quality is flat for the different tested connection configurations. The Paris average quality is the least as it has more information compared to the other tested videos, as noticed that its bitrate is near than the network BW. This leads to that none of the connection configurations reached the maximum quality. The figure also indicates that the (1,1) connection configuration is better.

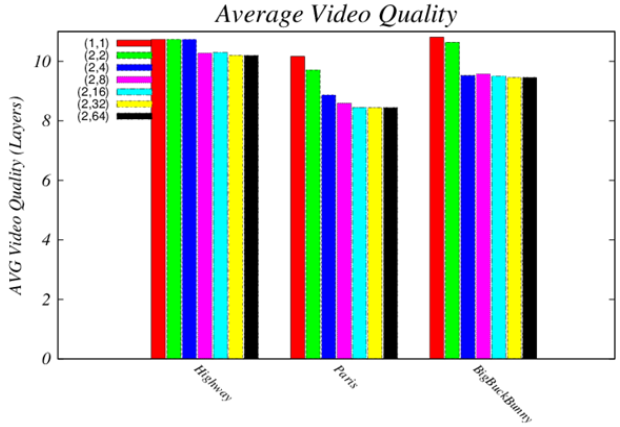
TABLE VI. INFORMATION ABOUT TESTED VIDEO SEQUENCES.

Video Name	Video size(MB)	Video duration(seconds)	Video Bit Rate(Mbps)
Highway	4.03	83.33	0.5
Paris	7.28	44.375	1.376
Big Buck Bunny	10.2	120	0.713



Average Video quality for different connection configurations (n_c^{min} , n_c^{max}) at 2 Mbps BW, and 10 ms link delay for the Horizontal Policy.

In order to study the relation between the quality and the PSNR. Figure 18 plots the average PSNR versus different videos considering different connection configurations for a BW of 2Mbps, and an average link delay of 10ms. The results confirmed the quality ones. The results show that the BigBuckBunny video has the highest PSNR, then Highway, and after that Paris.



Average Video quality for different connection configurations (n_c^{min} , n_c^{max}) at 2 Mbps BW, and 10 ms link delay for the Horizontal Policy.

We investigate the impact of both the horizontal and diagonal enhancement layer selection policies on our SVC-DASH-M streaming client when our client using the whole BW, and under real environment when there is cross traffics with the DASH client. The evaluation of the received video is done based on the video quality and the PSNR. The video quality is measured as the number of layers achieved. Experiments are done under several network conditions for different connection configurations when the video is streamed. Table VII provides a performance and implementation comparison between our SVC-DASH-M implemented system and the systems implemented in [25, 26]. It is worth mentioning that the proposed SVC-DASH-M algorithm is faster that than the systems in [25, 26] since a parallel segment fetching method is used. This is very important not to have any delay in the transmitted streaming video sequences. Additionally, the proposed SVC-DASH-M algorithm is implemented on a real testbed compared to the systems proposed on [25, 26]. It is worth mentioning that the proposed SVC-DASH-M system is evaluated using many parameters that accurately evaluate our system and measure its performance compared to those in [25, 26].

TABLE VII. PERFORMANCE AND IMPLEMENTATION COMPARISON.

	[26]	[25]	SVC-DASH-M
Protocol Type	**Application layer**	**Application layer**	**Application layer**
Video Quality Measurement in terms of:	- Bit rate in Kbit/sec. - The PSNR is not calculated.	- The received number of layers, The average is 8.5 at 10MB and 2 ms delay. - The PSNR is not calculated.	- The received number of layers. The average is 3.5 at 2Mbps and 2ms delay at the UDP cross traffic of 4Mbps. - The PSNR is calculated.
Adaptation technique based on:	Client-Buffer Status	Ratio between segment duration and its fetch time	Current Observed BW and Client-Buffer Status
The performance parameters	Buffered media time, and rate adaptation speed versus Competitive Bit Rate (CBR).	Buffer leveland throughput.	Number of interrupts, Duration of each interrupt, Number of opened connections, and Application goodput.
segment fetching method	Serial	serial	parallel
Evaluation	Implementation of a platform independent library integrated in a DASH client prototype.	Simulation on NS2	Implementation using Real Testbed
Algorithm behavior in presence of cross traffic	Stable	Stable	Stable

4. CONCLUSION AND FUTURE WORK

An adaptive SVC-DASH-M system was implemented using a real testbed over dynamic multiple connections. Improving the streaming quality at high buffer levels through downloading enhancement (successive) layers is the main advantage of our DASH client algorithm. Additionally, in case of streaming interruptions, our algorithm maintains the user experience quality at low buffer-level by focusing on the base layer segments. The dynamics of the used connections are considered as an additional critical design advantage of the proposed SVC-DASH-M. The proposed algorithm is experimentally tested under different connection and link configurations. Our results show that the algorithm successfully achieves interruption free streaming under all the tested BandWidth and link configurations. Additionally, the usage of multiple connections results in noticeable improvements in the achieved streaming quality for large link delays.

ACKNOWLEDGMENT

The authors acknowledge the support of Zewail City of Science and Technology – Cairo – Egypt and Southern University and A&M College - Baton Rouge - USA for their support to finalize this work.

REFERENCES

[1] T. Stockhammer, "Dynamic adaptive streaming over HTTP --: standards and design principles," *MMSys '11 Proceedings of the second annual ACM conference on Multimedia systems* pp. 133-144, San Jose - CA, USA, 2011.

[2] J.-M. Liang, J.-J. Chen, P.-C. Hsieh, and Y.-C. Tseng, "Two-Phase Multicast DRX Scheduling for 3GPP LTE-Advanced Networks," IEEE Transactions on Mobile Computing, vol. 15, pp. 1839-1849, 2016.

[3] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," IEEE Communications Surveys & Tutorials, vol. 17, pp. 469-492, 2015.

[4] Z. Yan, J. Xue, and C. W. Chen, "Prius: Hybrid Edge Cloud and Client Adaptation for HTTP Adaptive Streaming in Cellular Networks," IEEE transactions on circuits and systems for video technology, vol. 27, pp. 209-222, 2017.

[5] Y. Sánchez de la Fuente, T. Schierl, C. Hellge, T. Wiegand, D. Hong, D. De Vleeschauwer, W. Van Leekwijck, and Y. Le Louédec, "iDASH: improved dynamic adaptive streaming over HTTP using scalable video coding," Proceedings of the second annual ACM conference on Multimedia systems, pp. 257-264, 2011.

[6] F.-Y. Shih, C.-L. Fan, P.-C. Wang, and C.-H. Hsu, "A Scalable Video Conferencing System Using Cached Facial Expressions," International Conference on Multimedia Modeling, pp. 37-49, 2017.

[7] D. Gao, H. Lin, Y. Liu, and A. Jiang, "Minimizing End-to-End Delay Routing Protocol for Rechargeable Wireless Sensor Networks," Adhoc & Sensor Wireless Networks, vol. 34, 2016.

[8] A. Golechha, S. Karanje, and J. Abraham, "Comparative study of multicasting protocols based on average end-to-end delay," International Conference on Computing, Analytics and Security Trends (CAST), pp. 58-61, 2016.

[9] C.-H. Wang and T. Javidi, "Adaptive Policies for Scheduling With Reconfiguration Delay: An End-to-End Solution for All-Optical Data Centers," IEEE/ACM Transactions on Networking, 2017.

[10] J. Famaey, "On the merits of SVC-based HTTP Adaptive Streaming," IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp. 419-426, Ghent, 2013

[11] A. N. Park and W. Wei, "Adaptive streaming for digital content distribution," ed: Google Patents, 2017.

[12] S. Ibrahim, A. H. Zahran, and M. H. Ismail, "SVC-DASH-M: Scalable video coding dynamic adaptive streaming over HTTP using multiple connections," 2014 21st International Conference on Telecommunications (ICT), pp. 400-404, 2014.

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," IEEE transactions on circuits and systems for video technology, vol. 17, pp. 1103-1120, 2007.

[14] R. Deng and G. Liu, "QoE driven cross-layer scheme for DASH-based scalable video transmission over LTE," Multimedia Tools and Applications, pp. 1-25, 2017.

[15] Z. Ye, F. De Pellegrini, R. El-Azouzi, L. Maggi, and T. Jimenez, "Quality-Aware DASH Video Caching Schemes at Mobile Edge," 2017 29th International Teletraffic Congress (ITC 29), vol. 1, pp. 205-213, 2017.

[16] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, Y. Guo, and S. Cheng, "QoE-driven optimization for cloud-assisted DASH-based scalable interactive multiview video streaming over wireless network," Signal Processing: Image Communication, vol. 57, pp. 157-172, 2017.

[17] X. Qiu, H. Liu, D. Li, S. Zhang, D. Ghosal, and B. Mukherjee, "Optimizing http-based adaptive video streaming for wireless access networks," 2010 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), pp. 838-845, 2010.

[18] S. Lederer, C. Müller, and C. Timmerer, "Dynamic adaptive streaming over HTTP dataset," Proceedings of the 3rd Multimedia Systems Conference, pp. 89-94, 2012.

[19] N. Bouten, S. Latré, J. Famaey, F. De Turck, and W. Van Leekwijck, "Minimizing the impact of delay on live SVC-based HTTP adaptive streaming services," 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp. 1399-1404, 2013.

[20] T. Andelin, V. Chetty, D. Harbaugh, S. Warnick, and D. Zappala, "Quality selection for dynamic adaptive streaming over HTTP with scalable video coding," Proceedings of the 3rd Multimedia Systems Conference, pp. 149-154, 2012.

[21] C. Liu, I. Bouazizi, M. M. Hannuksela, and M. Gabbouj, "Rate adaptation for dynamic adaptive streaming over HTTP in content distribution network," Signal Processing: Image Communication, vol. 27, pp. 288-311, 2012.

[22] "JSVM BitStream Extractor library," Retrieved from http://ube.ege.edu.tr/~boztok/JSVM/SoftwareManual.pdf, Dec. 19, 2017.

[23] "Ubuntu 12.04 LTS "Retrieved from http://releases.ubuntu.com/12.04/, Dec. 19, 2017.

[24] "Arizona State University video library," Retrieved from http://trace.eas.asu.edu/yuv/, Dec. 19, 2017.

[25] C. Liu, I. Bouazizi, and M. Gabbouj, "Rate adaptation for adaptive HTTP streaming," Proceedings of the second annual ACM conference on Multimedia systems, pp. 169-174, 2011.

[26] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation algorithm for adaptive streaming over HTTP," 2012 19th International Packet Video Workshop (PV), pp. 173-178, 2012.

Ms. Samar Ali is currently a Teaching Assistant (TA) at Zewail City of Science and Technology - University of Science and Technology - Zewail City - Egypt. She also worked at Nile University, El Gezira High Institute, and Cairo University - Egypt as a TA. She previously worked as a Research Assistant (RA) at Media Streaming Optimization in Heterogeneous Wireless Networks Project in the Faculty of Engineering - Cairo University - Egypt. She received her BSc and MSc in Electrical Engineering from Electronics and Electrical Communication Department - Faculty of Engineering - Cairo University - Egypt in 2011 and 2016 respectively. Ms. Samar's research area of interest includes video streaming protocols and communication systems.

Dr. Yasser Ismail received his B.Sc. degree in Electronics & Communications Engineering from Mansoura University - Egypt, in 1999. He received his M.Sc. in Electrical Communications from Mansoura University - Egypt, in 2002. Dr. Ismail received his M.S. and Ph.D. degrees in Computer Engineering from University of Louisiana at Lafayette - USA in 2007 and 2010 and subsequently joined Umm Al-Qura University - Kingdom of Saudi Arabia as an assistant professor. In Fall 2012, he joined University of Bahrain - Kingdom of Bahrain as a Computer Engineering Assistant Professor. In Fall 2016, Dr. Ismail Joined both Electronics & Communications Engineering Department - Mansoura University - Egypt and Zewail City of Science and Technology - University of Science and Technology - Zewail City - Egypt as an assistant professor. Dr. Ismail is currently working as an assistant professor in the Electrical Engineering Department, Southern University and A&M College - Baton Rouge - Louisiana - USA. His area of expertise is Digital Video Processing Algorithms/Architectures levels, Internet of Things (IoT), VLSI and FPGA Design (Low-Power and High Speed Performance Embedded Systems), automotive transportation, Robotics, RFID, and Wireless and Digital Communication Systems. He has published one book, two book chapters, and more than 35 articles in related journals and conferences. Dr. Ismail served as a reviewer for several conferences and journals, including IEEE ICIP, IEEE GCCCE, IEEE ICECS, IEEE MWSCAS, IEEE ISCAS, IEEE SIPS, IJCDS, Springer, Elsevier, IEEE Transactions on VLSI, IEEE Transaction on Circuit and System for Video Technology (TCSVT), and IEEE Transactions on Image Processing. He severed in the technical committees of IEEE ISCAS 2007, IEEE ICECS 2013, MobiApps 2016, and IEEE MWSCAS 2018 conferences. He also served in the organizing committee of ICECS2013. He was invited to serve as a lead guest editor for special issue in mobile information systems – Hindawi publishing corporation September 2016. Dr. Ismail has been appointed as a Track Chair in the International IEEE Midwest Symposium on Circuits and Systems (MWSCAS) 2018. Dr. Ismail served as a PI and Co-PI for several funded grants from NSF and other international fund agencies. Additionally, Dr. Ismail served as a member in the Editorial Board Member for Frontiers of Mechatronical Engineering(FME, EnPress Publisher Editorial) - USA, 2018.

Dr. Ashraf Badawiis currently the Dean of Student Affairs at Zewail City of Science and Technology, Egypt. He is also the director for the Center of Learning Technologies. Dr. Badawi's conducts research in Educational Technology, Nanotechnology and Wireless Communications as a senior research scientist. Dr. Badawi was leading a team of research scientists at SMART Technologies exploring areas of research and development that is related to technology for the education and enterprise domains. In particular the current endeavors he is participating in are trying to define technologies to facilitate knowledge focused collaboration, to serve the education and enterprise markets in the fields of knowledge management, summarization, transcription, and student assessment. Prior to joining SMART, Dr. Badawi was the lead WiMAX Solutions Specialists for Intel in the Middle East and Africa where he was the technical lead for the Center of Excellence for Wireless Applications, a joint project between Intel and King Abdul-Aziz City for Science and Technology in Riyadh. Dr. Badawi was the systems architect for emerging market solutions from Intel, he moved back to Egypt in 2005 to be part of the Cairo Platform Definition Center. The team worked on the Classmate PC project that was later deployed to millions of students worldwide. Dr. Badawi graduated from the Systems and Biomedical Engineering Department in 1990 in Cairo, where he started pursuing his MSc in Engineering Physics. He earned his PhD for his work on the numerical analysis for microstrip antennas and high speed printed circuits in 2002. Dr. Badawi did his MBA in Marketing at the AUC, Cairo, graduating in 2009. He has more than 50 Journal, International Conference research papers.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.4 (July-2018)
http://dx.doi.org/10.12785/ijcds/070402

User Quality of Experience (QoE) prediction in Heterogeneous Mobile Networks

Mbemba HYDARA[1], Ahmed Dooguy KORA[2], Didier YANKAM[2] and Antoine GNANSOUNOU[2]

[1] *Department of Mathematics and Computer Science, University of Gaston Berger, St.Louis, Senegal*
[2] *Ecole Superieure Multinationale de Telecommunications (ESMT), Dakar, Senegal*

Received 14 Dec.2017, Revised 13 Mar. 2018, Accepted 25 Apr. 2018, Published 1 July 2018

Abstract: Quality of experience (QoE) is a key indicator in assessing the level of satisfaction of services offered by network operators. Mobile operators use several generation of technologies; these are not only based on the dynamic evolution of the networks but also the need to satisfy user expectations under intense competition. This study propose a methodology that will give subjective rating of communication taking place using two different technologies based on key performance indicators (KPI) parameters of 3G and 4G networks generations. Our approach goes beyond simple initiation of communication maintained by the same technology but involves several communications scenarios, which for various reasons may require switching from one generation of mobile network to another. In our proposed model, the ability to process survey results from a user QoE evaluation perspective is validated by our selection and use of appropriate combination of KPIs for each technology. The formula used will allow operators to predict duration that an Inter-Radio Access Technology (RAT) communication must have in each generation of mobile network services. This may help maximize customer loyalty and satisfaction based on selection of specific parameters.

Keywords: Heterogeneous,Mobile,Networks,QoE,Vertical handoff,Assessment

1. INTRODUCTION

The 1990s marked the beginning of real revolution in the field of telecommunications; an unbridled development which had the effect of changing the way of life of society at the same time giving rise to new needs. The emergence of several technologies and their deployment ushered in an increase demand on services.

While one technology is being deploy another is announced. Inevitably, operators have no options but to provide high communication quality under more complex and dynamic networking conditions [1].These developments are evidence of the co-existence of the various network technologies such as (2G/3G/4G) with variable weighting by country and zones around the globe. A heterogeneous network consists of different Radio Access Technologies (RAT), which includes among others High-Speed-DownlinkpacketAccess (HSDPA), LongTermEvolution (LTE), Wireless Local Area -Network (LAN), and Worldwide Interoperability for Microwave Access (WiMAX) networks [2].

The interaction between these technologies sometimes comes with the same service offer using same terminal in a complementary manner. The complexity resulting from these technologies by extension have consequences on the engineering, operations, maintenance and services in general. The challenge is how to navigate between these technologies seamlessly.

This is typical of a situation were a subscriber terminal is subject to change of technology communication; or coverage of one of the technologies being affected by the physical position of the user. To achieve the best connectivity and quality of service (QoS), a handover process should execute seamlessly so that ongoing sessions can be maintained [3]. A handover is a process where an ongoing call or data session of the mobile user is transferred from current network to a new available network. It is categorized into horizontal (HHO) and vertical handovers (VHO).

The former is when a mobile user is switched between same Radio Access Technology (RAT) for example (Wi-Fi to Wi-Fi) symmetric and the latter when user is switched between different RATs such as (Wi-

E-mail address: Hmbemba@utg.edu.gm,Ahmed.kora@esmt.sn,,Yankamdidier@gmail.com, Antoine.gnansounou@esmt.sn

MAX to LTE) asymmetric [3],[4]. Today, commercial industry tools have capabilities that allow us to predict quality of listening; hence, the challenge to navigate seamlessly between these technologies is possible. These tools ensure accuracy between predicted values in objective ways and those actually perceived by (subjective) users. This approach is possible even in the case of Inter-Radio Access Technology (RAT) appeals. The quest for a unified theory to reconcile the two forms of quality assessment methods is yet to be fully addressed.

The relevance of our approach is demonstrated by simultaneous collection of subjective and objective data in an attempt to determine an objective formula. For example, the main user lambda should be able to realize his communication or connection with the best satisfaction. That is to say, be able to switch to the desired network at the right time to maintain better quality of service. The network operator should also be able to guarantee QoS delivery based on the terms and condition of the license agreement.

The responsibility of Regulatory control of quality of service to users' falls under the purview of Telecommunications Regulatory Authority in each country. It is enforced through periodic campaigns and the findings presented to operators for corrective action if and when necessary. Due to market competition and fear of sanctions, operators carry out similar campaigns and usually require financial and human resources. To measure quality of the network, several tools and techniques are available in the industry for use in both objective (measurement and monitoring) and subjective (evaluation campaigns) assessments. Each of these methods has advantages as well as disadvantages.

This article proposes a formula that will give subjective rating of communication taking place using two different technologies based on KPI parameters of 3G and 4G network services. Our approach takes into account objective and subjective evaluation methods. The rest of the paper is organized as follows: Section 2 explores related works; Section 3 discusses QoS/QoE assessment methods. Section 4 expands on the proposed bi-generational conversion approach. Section 5 analyzed the results of the simulation and finally Section 6 draws conclusion.

2. REVIEW OF RELATED WORKS

Mobile devices establish connection with the network via several telecommunications operators. Users on the other hand have expectations about services they receive from operators [5]. These expectations combined with additional factors determine 'users' quality of experience (QoE) of a given system or service. The term quality of experience is often misunderstood and narrowly associated with QoS [6]. ITU Rec.E.800 [7] defines QoS as 'the totality of characteristics of telecommunications services that bear on its ability to satisfy stated and implied needs of the combined effect of service expectations or experience of the user". From a service provider's perspective, a concept by which network parameters are define, measured and controlled to achieve a level of service satisfaction.

The European Telecommunication Standard Institute (ETSI) takes a similar approach to that of ITU's in their definitions, based on 1988 version of the E.800 REC [8]. The Internet Engineering Task Force (IETF) has even more than ITU and ETSI, taken a network centric view of QoS with the following definition [9], 'a set of service requirements to be met by the network while transporting a flow'. In this definition, there is no mention whatsoever of 'users'. Quality of Experience (QoE) in contrast, is defined as: 'the overall acceptability of delivered service as perceived subjectively by the end user' [10]. Researchers characterize the term QoE as a multi-dimensional construct with subjective and objective factors intertwined in the user interaction as associated with perception, emotion, behavior, need, context, system and networking [11]. The concept is widely accepted and influenced by both system users and context centric factors [12].

QoE modeling has important benefits, for example, because of its ability to measure and predict allows the possibility of moving from systems oriented quality evaluation methods into a more user centric approaches. Several studies on QoE have been published using different methods and techniques. For example, in classification and regression method, machine learning, data mining and statistical modeling algorithms have been employed for the prediction of QoE [13]. QoE models are limited to QoS parameters [14]. In [15], Wu et al. gave a comprehensive account of QoE modeling problem. The authors proposed a conceptual model using QoE and QoS constructs. In their method, the parameters considered are concentration, attention and technology acceptance. However, this model did not take into account other context parameters such as location of user, type of mobile device used, and time of the day etc.

On the contrary, Mitra et al. [16] argues that 'inclusion of several context parameters in a QoE model could lead to an increase in QoE measurement and prediction accuracy especially in users' real-life environments. Therefore, the major challenge for operators and now customers is no longer based on the notion of 'the network offering the best quality of service', but the one best perceived from the point of view of the customer with better quality of experience (QoE).For example, users often have expectations about services offered to them by different operators. If they (users) are not satisfied with their quality of experience, they may switch

to different operator or stop using a particular application all together.

Using subjective and objective tests, QoE measurement can be performed [17]. Subjective tests involve direct data collection from users in the form of user ratings. For a given communication situations, service prescriptions and levels of QoS, the goal is to provide objective and subjective measures of users experience [18]. Quality of experience (QoE) is comprehensively explained in ITU International Standards. The goal of measuring quality parameters in the next generation networks with their impact on QoE is featured under ITU-T SG-12. A detail methodology for conducting subjective tests is also captured in [19], where the method for subjective test is presented. It defines a methodology for measuring users QoE based on Mean Opinion Score (MOS) rating. MOS is widely used for subjective voice and video quality assessment where human test subjects, grade their overall experience on the Absolute Category Rating Scale (ACR).

3. QoE ASSESSMENT METHODS

Quality of experience (QoE) often emerges where quality of service is no longer sufficient. Due to difference in human perception, a user does not usually perceive a service in the same way as his peers. In [20], QoE is defined as 'the overall acceptability of an application or a service, as perceived subjectively by the end user'. This definition is considered in some quarters as incomplete hence various institutions made the attempts to close the gap. For example, the European Network on Quality of Experience in Multimedia Systems and Services (QUALINET) through its white paper on Quality of Experience [21] endeavor a more comprehensive definition of Quality of Experience. It refers the term as the 'the degree of pleasure or annoyance of a user with respect to an application or a service'. It is the result of the fulfillment of expectations with regards to utility and or enjoyment of the application or service in the light of its personality and its present state'. This definition tends to place the customer more central and closer to users perception of the offered service. To better evaluate this perception of the user, several methods have been developed and grouped into Subjective surveys and objective methods.

The commonly used test methods are conversational opinion tests and listening opinion tests.

A. Subjective Surveys

TABLE1. SUBJECTIVE ASSESSMENT METHODS

Conversation- Opinion Test	Goal to produce as far as possible condition of services perceived by users. It is carried out in the laboratory. Conditions before and after experiment must be recorded and correctly preserved.
Listening –Opinion Test	Slightly less realistic than previous. The recommended test method for listening tests is Absolute category rating (ACR)
Quantal Response Detectability Tests	Allow to evaluate the threshold values of certain quantities and the corresponding probabilities
Degradation Category Rating (DCR)	Compares the system to be measured with a high quality fixed reference and the degradation (from 'inaudible to very annoying") is noted on a five (5) point scale
Comparison Category Rating (CCR)	Variant of DCR method. Compares the system to be measured with a high quality fixed reference (in the case of CCR, with a scale that goes from "much better" to 'much worse")
Threshold Method	Performs direct comparison of the target system with a reference system, such as modulated noise reference apparatus (MNRU).

The first requires special provisions, hence the second method was preferred, 'listening opinion tests' [22]. In [23], there is a distinction between the two types of subjective experiences; Passive and Active. In the active or interactive experiments, at least two participants were engaged in a conversation using means available to them. In these cases, participants follow certain protocols in accordance with a set plan. A statistical sample of 100 participants were used; 50 males and 50 females young and old. In [24], the text to be pronounced for the recording must be short, simple and clear. They must be chosen in a random manner, with no relationship between them to allow the evaluator concentrate solely on the quality of what he or she perceives'.

In our study, we opted for a passive listening opinion tests. We conducted a conversation in an environment familiar to participants, while raising the key performance indicators of each conversation, and have listened to a panel that provided us with their feelings through the notation proposed to them. With the passive environment, their opinions were given based on the scale provided to them. The scoring of the conversations heard by the users was done on several different scales.

The absolute Category Rating (ACR), alternative Discontinuous Category Rating (DCR), the assessment by Comparison Category Rating (CCR) and the threshold method [25]. For our case, we applied a five (5) point ACR Scale (Absolute Category Rating) in TABLE 2 [26], [27].

TABLE 2 ACR RATING SCALE

Listening Quality Scale		
Score	**Quality**	**Disturbance**
5	Excellent	Inaudible
4	Good	Perceptible, but not disruptive
3	Fair	Moderately Disruptive
2	Poor	Disruptive
1	Bad	Very disturbing

B. Objective Method

In addition to subjective factors, which have the main disadvantage of being costly, other methods have been developed based on certain parameters. These methods are develop by entities such as ASCOM's with the Speech Quality Index (SQI), SEVANA's Passive Voice Quality Analysis (PVQA), or those that have been standardized by ITU POLQA (Perceptual Objective Listening Quality Analysis) [28], which replaced the Perceptual Evaluation of Speech Quality (PESQ) [29]. The advantages of POLQA with respect to PESQ that justify its replacement are as follows:

- Maintain good evaluation level despite background noise
- Equations with commas or periods takes into account speech level in samples (KHz)
- Sensitive to linear distortions
- Create new scale for SWB signals and SWB (48) and from our analysis, we retained various parameters for use in the implementation.
- Super Wideband (SWB) level from 50KHz to 14KHz of our formula
- Allows comparison between the AMR codec used in GSM/3G and the EVRC codec used in the CDMA 2000.
- Takes into account two different sampling frequencies depending on the band;NB(8KHz)

TABLE 3. FEATURES OF THE DIFFERENT METHODS

	POLQA	**PESQ**	**PVQA**
Operating mode	Defined two operating modes: ▪ Super Broadband mode with the following bandwidths: ➢ Super broadband ➢ broadband ➢ Narrow broadband ▪ Narrowband mode for narrowband networks.	Defines several versions in order to compare notes of different technologies: ➢ PESQ-Wide Band (WB) ➢ PESQ-Narrow Band (NB)	Uses two operating modes: ➢ Non-intrusive calculation of MOS ➢ Bulk Fault Detection throughout the audio test
Input parameters	They take as input at least three parameters: ➢ The original file as it should be issued (issuer registration) ➢ The "degraded" file that has already passed through a transmission system (recording what the receiver perceives) ➢ The sampling rate		The PVQA uses 6 input parameters: [24] ➢ "Pvqa.lic" which is a license file issued by Sevana. ➢ Analysis and / or "graph" are the parameter that defines the mode of operation of PVQA. ➢ "ENG_F_40.wav.csv" is the name of the report file where PVQA will store information about the alterations found in the defined slots. ➢ Settings.cfg is a PVQA parameter file prepared and provided by Sevana. ➢ ENG_F_40.wav is an uncompressed wav file for testing ➢ 0.799 is the time interval in seconds that the PVQA will use to analyse for depreciations on the one hand and then to predict the MOS score.

Parameters used in the algorithm	POLQA uses 6 parameters; ➢ a frequency response indicator (FREQ) ➢ a noise indicator (NOISE) ➢ a reverberant room indicator (REVERB) ➢ Three internal indicators; propagation time, the quantization step and a voice noise indicator	❖ The PESQ used three parameters, namely ➢ A propagation time indicator ➢ A distortion indicator due to coding ➢ An indicator of transmission error in the voice	Ownership Algorithms
Adjustment	The Root Mean Squared Error (RMSE) method.	The correlation coefficient (CC)	Reserve Owners
Chosen field	Used for the comparison of different networks: 3G and 4G networks ➢ VoIP and NGN networks offering HD quality voice services such as "broadband" and "super-broadband" phone calls, the 7 kHz and 14 kHz frequency range	Desirable for: ➢ Networks still using G.711 audio codecs, law a, law u. ➢ Networks with low bandwidth from 300 to 3400 Hz of bandwidth. ➢ Also supports the WB (frequency range ➢ 7 kHz) using PESQ ITU-T P.862.2.	Desirable for IP networks [25] subject to license.
Future for different standards	New standard in force It has a relatively fast operating capacity and is more accurate than the previous ones. It solved some problems inherent in previous versions.	It will still be used for a number of years because of its backward compatibility and because many countries still have narrowband networks.	-

Quality of Service (QoS) is one of the ingredients that advertisers sell to customers. Under the law, the duty of the regulator to monitor quality of service is an indication of "good health" of a network. ITU defines QoS as 'the ability of a network or part of the network to perform functions related to communication between users" [26]. During a call, the mobile phone exchanges data with the network. In the upstream direction, it is the results of measurements made by the mobile phone sent to the network.

In optimization standards, KPIs are grouped into five (5) distinct classes: Accessibility, Mobility, Integrity, Continuity and Availability [27],[28]. Between the parameters that mobile phone exchanges with the network, we chose 3 indicators: CPICH RSCP (Common Pilot Channel Received Signal Received Power), CPICH $^{Ec}/_{N0,}$ RSRP (Reference Signal Code Power), RSRQ and BLER. The RSRP and CPICH RSCP as their names indicate, these two indicators are different and show the level of power received from the pilot channels.

The Reference Signal Strength Indicator (RSSI) indicates power level over full length of the bandwidth. Whereas RSRP and the RSCP, indicate the level of attenuation undergone by the signal of a user with respect to the channel used. The channel here takes into account the useful signal, noise and interference. The RSRP (4G) represents the received power level of the user cell in a Radio Block (RB) and the RSCP (3G) represents the received power level of the pilot frequency [30].

These two indicators are good for our test because their comparisons help decision-making in cases of technology change (handover inter RAT). Although providing essential information, the two notions do not provide information on the quality of the link or connection. The reference Signal Received Quality (RSRQ) and the CPICH $^{Ec}/_{N0,}$ are ratios between the power of the received signal in the active cell (4G) RSRP or the (3G) RSCP and the other received signals (RSSI in both generations) that are considered to be noise. They are measured only when the mobile phone is in dedicated mode (in this case, it is a voice conversation). The measured RSRQ varies between (-19.5dB and -3dB in 0.5dB) steps. The term $Ec/_{N0,}$ is a composite term: the Ec represents the energy received by the chip and the N0 being the total noise. The image commonly used to describe this term is the estimate of the Signal-to-Noise ratio hence the following formula:

$$E_c/_{N_0} = \frac{RSCP}{RSSI}$$

The Block Error Rate (for both technologies) measures transmission errors and is therefore effective at the physical layer. The terminals must support the BLER measurement. Therefore, the main measurement function of the BLER is to provide feedback for the external loop power control operation. In order to control the power, the second shortest seconds are allocated to the remote user of the base station after the scrambling codes, knowing that theoretically one more chip corresponds to a distance of 70m (in 3G).

The BLER when at 4G, its normal conditions of usage are 2% in an inbound synchronization and 10% in outbound synchronization [31], [32].

4. QoE PREDICTION APPROACH FOR COMPLEX NETWORKS

Quality of experience prediction tool allows us to extract particular information from the file in question, and in our case are the parameters we need. The extraction software also allows us to save the information as an Excel file. The Excel file once embedded, the Mean Opinion Score (MOS) can be calculated using MATLAB line connection Toolbox to establish a predictive model. Besides, the Tems Discovery, other appropriate tools such as ATIX also have capability to process information. Tools such as Excel, Magic 3 were also used during the simulation.

2) In the second step, we created a montage using specialized software called MAGIC Music Editor 3 to achieve the goal of determining variation in duration and generation. For example, duration of 3 minutes etc. Different timings were set for each network. In carrying out the test, we used 200 inter – RAT calls. For each of these calls, we recorded the KPI parameters. As stated above, calls were made on both 3G and 4G. We then proposed establishing a formula in two levels. The first level, a score of the conversations on each generation and the second combined the two previous levels to give an overall rating. These two formulae were based on the experimental results we had before. As a reminder, we identified three parameters for each of the generations (3G, 4G); the following formulae were derived using statistical model in the excel tool.

- In 3G: CPICH RSCP, CPICH $^{Ec}/_{N0}$and BLER
- In 4G: RSRP, RSRQ and BLER

We realized that some of these parameters have a proven similarity. We can express the RSRP based on the RSRP. The RSRQ is given by the following formula:

$$RSRQ = 10\,x \log\left(N\,x\,\frac{RSRP}{RSSI}\right)$$

With N as number of RB

For the 3G, thanks to the use of software prediction, we expressed the BLER according to the CPICH $^{E_c}/_{N_0}$. In our evaluation of the case study, a network receiver can switch from 3G to 4G in vice versa during a call (vertical handover). However, where a caller does not change position using same technology no vertical handover takes place. Geographically transmitting and receiving are supposed to be vertical handover. In our case study, both the caller and receiver remain in the same mobile network enabling different technology (3G-4G); in which case horizontal handover does not apply. The following diagram gives us more precision;

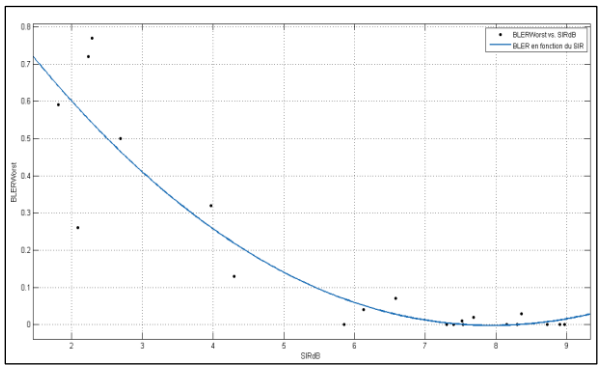


Figure 1. Prediction of the BLER according to the CPICH $^{E_c}/_{N_0}$

With the following features:

$$BLER = -0.0001932x^3 + 0.02053x^2 - 0.2894x\ 1.098$$

- With x$BLER = -0.0001932x3 + 0.02053x2 - 0.2894x\ 1.098$ - With x for CPICH $Ec\ N0\ /$

Goodness of fit:

- Adjusted R-square: 0.9001

Goodness of fit:

- **In 4G** : RSRQ and BLER
- **In 3G** : CPICH $^{E_c}/_{N_0}$ and BLER

$$F(x,y) = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}x.y + p_{02}y^2 + p_{30}x^3 + p_{21}x^2.y + p_{12}x.y^2 + p_{03}y^3$$

With

A. *4G-Formula*

- F(x, y) represent theMOS_{4G}
- x stand for RSRQ
- y stand for BLER
- $p_{00} = 0.009207\ (-0.09232, 0.1107)$
- $p_{10} = 0.02309\ (-0.09239, 0.1416)$
- $p_{01} = 0.08078\ (-0.09239, 0254)$
- $p_{20} = 0.04276\ (-0.07421, 0.1597)$
- $p_{11} = -0.02266\ (0.08342, 0.09985)$
- $p_{02} = 0.2011\ (0.08342, 0.3188)$
- $p_{30} = -0.0103\ (-0.07697, 0.05637)$
- $p_{21} = -0.1665\ (-0.3239, -0.009187)$
- $p_{12} = -0.07132\ (-0.2338, 0.09119)$
- $p_{03} = 0.07205\ (-0.03726, 0.1814)$

A. *3G-Formula*

- F(x,y) represent BLER
- x stand forMOS_{3G}

- y stand for CPICH RSCP
- $p_{00} = 0.02309\ (-0.1191, 0.1653)$
- $p_{10} = 0.07869\ (-0.1095, 0.2668)$
- $p_{01} = -0.118\ (-0.3382, 0.1021)$
- $p_{20} = -0.01746\ (-0.1141, 0.07923)$
- $p_{11} = -0.07472\ (-0.2962, 0.1467)$
- $p_{02} = 0.1687\ (0.07715, 0.2604)$
- $p_{30} = -0.01542\ (-0.07503, 0.04419)$
- $p_{21} = 0.07461\ (-0.161, 0.3103)$
- $p_{12} = 0.009085\ (-0.2753, 0.2935)$
- $p_{03} = -0.03633\ (-0.1236, 0.05097)$

B. General formula

$$MOS_{global} = \frac{\alpha}{\alpha + \beta}(MOS_{4G}) + \frac{\beta}{\alpha + \beta}(MOS_{3G})$$

Where

- αrepresents the duration of the communication in 4G
- β represents the duration of the communication in 3G
- MOS_{4G} represents the score that would be obtained if the conversation was only in 4G
- MOS_{3G} represents the score that would be obtained if the conversation was only in 3G

The HVCR formula acquired was validated when we compare between subjective values and the ones predicted.

TABLE 4. COMPARISONS OF SUBJECTIVE AND PREDICTED VALUES

Call Number	Subjective MOS	MOS Predicts	Percentage in 4G(%)	CPICH RSCP (dBm)	SIR (dB)	BLER (Worst - %)
1	4,166666667	3,315492958	85,91549296	-73,9	7,7	0
2	3,833333333	3,837837838	86,48648649	-66,4	8,3	0
3	3,333333333	2,996721311	83,60655738	-75	4	0,3
4	3,333333333	3,492307692	84,61538462	-77	6,2	0
5	3,166666667	3,4	75,6097561	-76,3	8,8	0
6	3,333333333	3,913043478	78,26086957	-74,4	8,9	0
7	3,833333333	3,497560976	75,6097561	-62,3	2,2	0,2
8	3,333333333	3,8	71,42857143	-69,5	8,2	0
9	4,166666667	2,981818182	84,84848485	-79,7	4,4	0,1

In order to appreciate the Global Mean Opinion Score (MOS3g/4g) for the duration of communication of 3G and 4G technologies based on the rating for the communication, we presented our approach and results in the following section.

C. *Approach*

In our calculation of MOS (3G-4G), we consider different cases of time allocation of the respective technologies in order to simulate the global MOS of the subscribers. This is very important since it could help predict appropriate quality allocation based on the technology while proceeding to a handover. The simple case is while both technologies (3G, 4G) are allocated equal times during the communication, the following expression is used;

$$\alpha/(\alpha + \beta) = 0.5 \quad (1)$$
$$\beta/(\alpha + \beta) = 0.5 \quad (2)$$

The second case corresponds to y when 4G is allocated three quarter of the time slot and x when 3G is allocated quarter of the time. This can be expressed as follows:

$$\alpha/(\alpha + \beta) = 0.75 \quad (3)$$
$$\beta/(\alpha + \beta) = 0.25 \quad (4)$$

The third case correspond to y when 4G is allocated a shorter duration (one fifth) of communication and 3G a longer duration (four fifth). These cases are respectively simulated in figures 2, 3, 4 in accordance with clause A.4.5.ITU P.800 standard. According to this clause, graphs should be plotted showing Mean Opinion Score (MOS) as a function of the parameters under test'. Our test results show the vertical axis Z as in figures 2, 3, 4 represents the mean opinion score (MOS) of 3G-4G as depicted in 3D graphic plots. The results of the survey demonstrate that being able to predict duration of time for a complete conversation using 3G/4G helps improve quality of satisfaction of users.

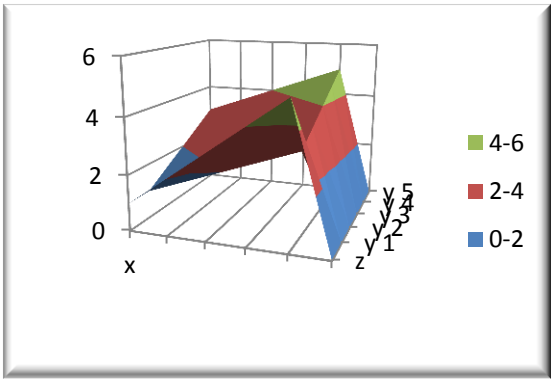


Figure 2. Graphic plots of global $MOS_{3G/4G}$Communication

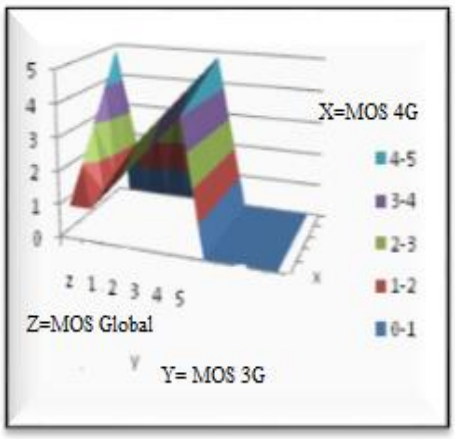


Figure 3. Graphic plots of global $MOS_{3G/4G}$ Communication

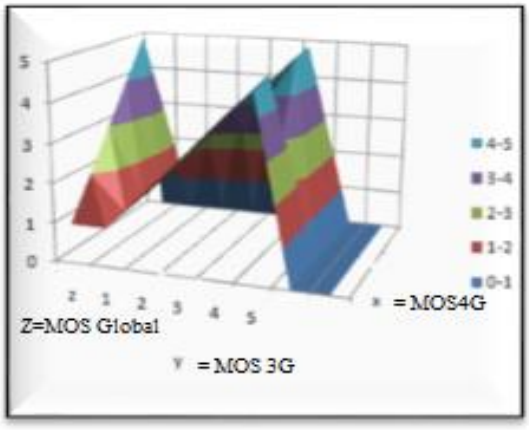


Figure 4. Graphic plots of global $MOS_{3G/4G}$Communication

E. Analysis of figures 2, 3, 4

With the foregoing, the different cases of time allocations of the respective technologies (3g, 4g) simulate the mean opinion scores (MOS) depicted in Fig. 2,3,4 graphic plots. In Fig. 2 for example, when value of x and y = 1 or 2, the communication link or connection is considered unacceptable. If equal 3 or 4 and $y \geq 3$ or y = 3 or 4 and $x \geq 3$. The communication is still considered poor. If x= 4 and y = 2 or above or y = 4 and x at least 2 or above, the communication is considered good. This is true because the quality of the communication is excellent at least when half of the time is utilised. To be acceptable, the communication should last at least half of the time. However, it should also be noted that while the communication might be excellent for 4G, it may be poor for 3G.Fig. 3 illustrate any value of x when y is 1, the communication is considered unacceptable.
Even if y is 2, it is still considered bad. The communication of both x and y can only be acceptable when x and y values are 3 or above.

In Fig. 4 when y = 1 indicate global quality is bad for any quality of 3G during the communications. The global quality is still bad for y = 2 in particular for x = 1. For better rating of 3G ($x \geq 2$) with y = 2 (MOS) during the same communication leads to a global poor quality of the communication. The communication becomes acceptable only when y and x are 3 or above.

5. Conclusion

The major challenge for operators and now customers is no longer based on the notion of "the network offering the best quality of service', but the one best perceived from the point of view of the customer with better quality of experience. Through this study, we have produced a model that will allow operators to better estimate or predict the duration that an inter-Radio Access Technology (RAT) communication must have in each generation of mobile technology. This will help maximize customer satisfaction at the same time increase loyalty based on specific parameters.

References

[1] A. Gotchev. K. Mueller. G. BozdagiAkar, D. Strohmeier, A. Boev, A. Tikanmäki, A., and Oksanen "MOBILE3DTV: State of the art of technology and standards, Project No. 216503; 2011, pp 1-30

[2] W. Abdellaziz, K Abdellatif, M.Abdelfettah et al. 'Group vertical handoff management in heterogeneous network: W*ireless Communications and Mobile Computing*, 2015, Wiley Online Library (wileyonlinelibrary.com).DOI: 10.1002/wcm.2599

[3] l. Joha, k. Loveneet, S. Sandhu Amandeep "An overview of Vertical Handover process and Techniques, *Indian Journal of Science and Technology, (IJST)* Vol.9 (14), April 2016 DOI: 10.17485/ijst/2016/v914/86601

[4] l. Joha, k. Loveneet, S Sandhu Amandeep, "An overview of Vertical Handover process and Techniques, *Indian Journal of Science and Technology, (IJST)* Vol.9 (14), April 2016 DOI: 10.17485/ijst/2016/v914/86601, 2016

[5] R. Schatz & Hossfeld, Tobias & Janowski, Lucjan & Egger-Lampl, From Packets to People, "Quality of Experience as a New Measurement Challenge: *Telecommunications Research Center (FTW), Vienna, Austria* 219263 Sebastian (2013) 10.1007/978-3-642-36784-7_10

[6] Waleed, N. Mahmoud, S. Mona, "Multi-Access Interference Mitigation Using Multi-Level Power Control Algorithm in OFDMA Cellular System, *IJCDS*, Vol. 5, No.6, Nov. 2016

pp. 1-7

[7] P. Le, Callet, S. Moller, and A. Perkis (Eds), "Qualinet White Paper on Definitions of Quality of Experience, 2012, Lausanne, Switzerland

[8] R. Jain, "Quality of experience Multimedia, IEEE, Nov 11, 2004

[9] ETSI" ETR 003 Network aspects; general aspects of quality of service (QoS) and network performance (NP), 1994

[10] E. Crawley, H. Sandick, R. Nair, B. Rajagopalan, 'A framework for QoS based routing in internet, IETF RFC 2386, 1998

[11] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. 'Yang, "Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. *In conf. of the seventeen (17) ACM international conference on Multimedia*, pages 481–490, New York, NY, USA, 2009

[12] Sun J. 'Football on mobile phones: algorithms, architectures and quality of experience in streaming video. *PhD. Thesis*, Umeå University, 2006

[13] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt., R. Sheppard, and Z. Yang: 'Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. *In Conf. on Multimedia of the seventeen (17) ACM international conference*, pages 486–490, New York, NY, USA, 2009.

[14] A. Bhattacharya. W. Wu., and Z. Yang, "Quality of Experience Evaluation of Voice Communication: An Affect-Based Approach, *Human-Centric Computing and Information. Sciences*, Vol. 2, issue 1, 2012 pp. 1–18.

[15] T. Hossfeld, S. Kapov Lea., E. Heegaard Poul. M. Varela., T. Chen Kuan., 'On Additive and Multiplicative QoS-QoE models for Multiple QoS Parameters, Workshop report on Perceptual Quality of Systems, Aug 2016, pp.29-32, Berlin, Germany

[16] K. Mitra., C. Åhlund, A. Zaslavsky,"A decision-theoretic approach for quality of experience measurement and prediction, *In Conf. Multimedia and Expo (ICME*), *IEEE*, July 2011, pp 1–4,

[17] Methods for subjective determination: ITU Recommendation P.800

[18] Vocabulary for performance and quality of service Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100 (2008)

[19] Methods for Subjective determination of transmission quality. 1996 - ITU-T Recommendation P.800:

[20] Methods for Subjective determination of transmission quality 1996, Recommendation P.800.ITU-T

[21] P. Le Callet., S. Moller., A. Perkis. (Eds) "Qualinet White Paper on Definitions of Quality of Experience (2012)", Lausanne, Switzerland

[22] ITU Mean Opinion Score interpretation and reporting, Recommendation P.800.2, 05/2013

[23] ITU Méthodes d'évaluation subjective de la qualité de transmission, Recommandation P.800, 08/1996

[24] ITU-T Recommendation P.800: Methods for Subjective determination of transmission quality. 1996

[25] ITU-T, Methods for subjective determination Recommendation P.800.

[26] Perceptual objective listening quality assessment, Recommandations ITU, P.863, 09/2014

[27] I. Nafiz Bin Hamid1, L.Yahia, H. Mugumya Twarik, S. Nafiu "Towards an Efficient Radio Network Planning of LTE and Beyond in Densely Populated Urban Areas, IJCDS, 4, No.2 (Apr-2015) http://dx.doi.org/10.12785/ijcds/040205

[28] AB Abdullah, R Vaidyanathan, E Ariwa, W El-Medany 'Introduction to Special Issue: Design and Performance of Networks on Chip' *ICJDC, Vol.5 (2)* 2015

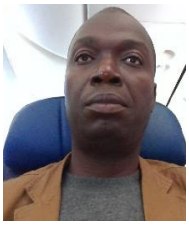
[29] Www. Sevana, Biz, accessed date 20/01/2017

[30] ITU, 'Terms and Definition of quality of service and network performance including dependability, 08/1994.26/08/2016

[31] H.Tobias, H.Poul E, V.Martin,'QoE beyond the MOS: Added value using quantiles and distributions, in Conference, 2015, Doi-10.1109/QoMEx. (2015.714142) pp.1-7

[32] Www. Tech-invite.com, Accessed date 27/08/2016

Mbemba HYDARA is a PhD student in Computer Science with the University of Gaston Berger- Senegal. In 1998 – 2000 he graduated with Advanced Diploma in Telecommunications and Networking. In 2003, he received a Masters' Degree in International Law with the University of Derby in the United Kingdom (UK) and in 2010 received a second Master of Science (MSc.) Degree in Forensics Computing & Security. He is IRCA- ISO27001 - ISMS Auditor/Lead Auditor registered. His current research is in the area of Telecommunications Quality Audit & Security.



Ahmed D. KORA is a graduate in Physics Sciences in 1998 from "Faculté des Sciences Techniques" at "Université d'Abomey–Calavi", Bénin, where he got his Diplôme d'Etude Approfondie (DEA) in Material Sciences in 2000. In 2003, he received a Master "Réseaux Télécoms" degree from "Ecole Superieure Multinationale de Telecommunications" (ESMT) and Ph.D. degree in telecommunication from the University Of Limoges, France, in 2007. He is currently with the ESMT and Head of Research and Innovation Department. His research area covers communications, radio and optical networks system architecture, universal access, mobile network quality of service and quality of experience, low cost IT systems for development, etc. Prof. KORA is IEEE member and member of Fiber Optic Association.

YANKAM is a graduate in Telecommunications and Networks from Ecole Superieure Multinationale des Telecommunications (ESMT) Dakar-Senegal where he received a Master's degree in 2016.

In 2012; he finish his first degree in Telecommunications Engineering with SUP'PTIC ''Ecole Nationale Superieure des Postes, des Telecommunications et des Technologies de l'Information et de la Communication" Yaoundé-Cameroon. He is also a Cisco Certified Network Associated in Routing and switching (CCNA).

Antoine GNANSOUNOU is a graduate in Mathematics from "Faculté des Sciences techniques" at "Université d'Abomey - Calavi", Benin where he received his "Maitrise ès Sciences Mathématiques" in 1991. In 2000, he received a "Réseaux Télécoms" master degree from "Ecole Superieure Multinationale des Telecommunications" (ESMT) and in 2007 a master of Research in complex systems simulation, Telecommunications at "Université Cheick Anta Diop de Dakar (UCAD)".

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.4 (July-2018)

http://dx.doi.org/10.12785/ijcds/070403

Designing Non contact based ECG System for Driver Drowsiness Detection

Srihitha Jujhavarapu[1], Mohsen Babaeian[1] and Mohammad Mozumdar[1]

[1]*Department of Electrical Engineering, California State University, Long Beach, USA*

Received 20 Dec. 2017, Revised 2 Feb. 2018, Accepted 9 Mar. 2018, Published 1 July 2018

Abstract: Driver drowsiness has been a significant hazard resulting in various traffic accidents, severe injuries, or even fatalities. Therefore, monitoring this condition is crucial not only in alerting drivers but also in avoiding fatal accidents. Therefore, this paper proposes a hardware design for drowsiness detection; in addition, the outputs used to justify this paper were simulated in the LT Spice. Through a thorough observation, it is apparent that a driver's drowsiness is associated with an immediate change in his heart rate, and due to the fact that Electrocardiogram (ECG) is used to detect an accurate heart rate, we used it as a parameter in the proposed design where it consists of a non contact ECG sensor as an input source and a circuit with a two-stage amplifier to improve the ECG signal's strength and filters to minimize noise. An approximate maximum peak ECG output voltage of 2.8V was obtained in LT Spice, and the resulting ECG output is sufficient enough to detect a driver's drowsiness while preventing major accidents.

Keywords: Electrocardiogram (ECG), Drowsiness, LT Spice.

1. INTRODUCTION

Recently, the rate of traffic accidents has increased in conjunction with the increase in vehicle numbers. Among the array of diverse automobile accident scenarios, driver drowsiness is one of the most dangerous situations; similar to alcohol or drugs, driver drowsiness, which can be caused by fatigue, sleep deprivation, extensive driving, a low circadian rhythm, or medication use, can be detrimental to the human brain. According to World Health Organization (WHO), over 1.5 million people die per year and over 40 million people have severe injuries resulting from driver drowsiness related accidents [1]. Due to this, there is a high need for developing a system that detects drowsiness and alerts drivers against hazards.

A driver's drowsiness is detected through various approaches such as analyzing a driver's physical behavior, vehicle response, brain waves, pulse rate, and respiration, but currently, detection using heart waves (e.g. Electrocardiogram (ECG) signals) is one of the most interesting topics of driver drowsiness detection [2]. Since drowsiness is associated with sleep, these physiological measurements provide accurate results that are based on the correlation between physiological signals and sleep. Because the driver's body automatically generates physiological signals and he/she has no control to alter them, the physiological signal detection methods have certain advantages over body movement pattern detection techniques such as eyelid movement. [3] Among the array of physiological signal analysis methods used for detecting driver drowsiness, the ECG analysis method proves more proficient due to its reliance on Heart Rate (HR), and because of this, it has been used in the proposed design.

The system's design begins with a model and simulation in order to validate the results before implementing them in real-time. Electronic circuit simulators such as LabVIEW, PSPICE, MULTISIM, and LT Spice are the various tools utilized to implement the circuit designs, but due to its feasibility and ease of design, LT Spice is the main tool used for simulation. In addition, the ECG signal comprises a human system's physiological information and is used as an input for this paper's LT Spice simulation. In conclusion, the simulation results that were obtained are adequate enough to successfully implement the proposed drowsiness detection design in real-time.

The remaining portion of this paper is methodized as follows: Section 2 reviews related work, Section 3 provides background information on the electrical heart

E-mail: srihitha.jujhavarapu@student.csulb.edu, mohsen.babaeian@cgu.edu, mohammad.mozumdar@csulb.edu

signal and its analysis, Section 4 explains the methodology used in our research and the circuit simulation, and this paper concludes by discussing the overall main points in Section 5.

2. RELATED WORK

Due to various driving hazards including driver drowsiness, driver safety has been one of the leading challenges faced by the automobile industry. However, due to the prevalence of driver drowsiness related incidents, several methods have been proposed for drowsiness detection throughout the years. Sahayadhas [4] suggested a hybrid driver drowsiness detection system that utilizes multiple sensors (i.e. a strain gauge sensor and ECG sensors). In his research, he proposed two methods to determine a driver's drowsiness Firstly, it is detected using a verbal questionnaire; in addition, drowsiness levels are evaluated using the driver's response and several tools have been used to convert this rating to measure a driver's drowsiness. Secondly, by utilizing sensors, the change in the data is analyzed to measure drowsiness. The limitation of this system is verbal questionnaire cannot be implemented during driving [5]. Swapnil [6] suggested a driver's fatigue detection technique using a strain gauge sensor; this system consists of a sensor, a signal processing module, and an alarm-producing device. The sensor is positioned in front of the driver and monitors eye and jaw movements micro-sleeps and obtains input data. As soon as the driver feels drowsy, fatigue is detected, and the system produces an alarm to alert the driver, but while this technique is innocuous, the design is costly and requires complex processing techniques. Ghosh [7] presented a computer vision system that uses eye tracking to monitor driver drowsiness in real-time; the system consists of a camera that captures images of the face, a data acquisition block that implements algorithms for face, eye, and pupil detection, and a processor that analyzes drowsiness from eye movements. However, the main drawback of this method is a change in the eye's sensitivity, which produces false positives in the results. Assari [8] proposed a system consisting of a camera and an infrared LED, and by using infrared LEDs, it was able to overcome the issue of intensity variation that resulted from external car lights. Nevertheless, the main shortcoming of this technique is that it yields fault results; for instance, the system will determine a person is drowsy from his appearance, when, in fact, he may not be.

Kumar [9] suggested a drowsiness detection system using electroencephalogram (EEG) signals; it consists of an EEG detection circuit, which comprises an EEG sensor that detects and amplifies the tiny electrical voltages generated by brain cells, a micro-control unit that produces a control signal for processing detected EEG signals, and an EEG signal processing circuit that then processes the EEG signals to identify the driver's drowsiness. The EEG sensors used in this design are wired; hence, it is not suggestible that they be used while driving because it impairs driver's concentration. Deepa [10] proposed a system for drowsiness detection using EEG signals. It contains an EEG sensor that detects EEG signals from the brain, an EEG signal acquisition unit, which amplifies and filters the signal for drowsiness analysis, and a mobile unit that alerts the driver if drowsiness related information is observed in the measured EEG signals. Though, it is necessary to note that the wired EEG sensors used in this proposed system are not safe to implement as it affects the driver's concentration.

Sang-Joong Jung [11] presented a new ECG sensor with conductive fabric electrodes, which can be positioned on a car's steering wheel in order to identify a driver's drowsiness, and the measured ECG signal from the driver's palm has a sampling rate of 100HZ. Yet, the sensors utilized in this design are complicated to employ because the skin-electrode impedance may result in poor quality of the ECG signal. Xun Yu [12] suggested a driver's drowsiness system using two non-intrusive conductive fabric ECG sensors placed on the steering wheel and the back of the driver's seat. The design consists of signal conditioning circuitry such as a notch filter, a bandpass filter, an amplifier and a driven right-hand circuit for improving the strength of the ECG signal obtained from two different sensors. In his proposal, he implemented an adaptive filter algorithm in the software to reduce the baseline noise. Though, it is necessary to note that sensors used in this method will fail to sense the ECG data if a person is wearing any gloves or if the driver uses only one hand. Sangeetha [13] presented an embedded driver drowsiness detection system that not only monitored and controlled the drowsiness state but also provided feedback to stop the automobile once drowsiness was identified. The input obtained from an ECG sensor is amplified and processed to determine the driver's state, and if the driver is drowsy, the processor notifies the driver by emitting an alarm and activating the driver circuit. However, because the design has electrodes fixed to thumb, this leads to impaired driving and therefore, not recommended.

Based on the above-related work, this paper proposes a design that is explained as follows:

- The drowsiness detection system is designed using non contact sensors.
- A low pass filter with a cutoff frequency of 33 HZ is used at each sensor to remove artifact and electrode contact noises.
- Two amplifier stages are utilized, strengthening the ECG signal, and minimizing common mode interference.
- A twin T notch filter with a cutoff frequency 60 HZ is used to remove the powerline noise.

- Simulation is performed in LT Spice with ECG signals as the input source.

3. BACKGROUND

A. *Electrocardiogram (ECG)*

ECG is a graphical depiction of electrical activity produced by the human heart via the utilization of electrodes placed on skin; these electrodes identify the small electrical variations on the skin that generate from the heart muscle's depolarization and repolarization patterns during each beat.

The typical ECG is made up of PQRST and occasionally U wave. Each PQRST waveform symbolizes a single heartbeat. One cycle of an ECG wave is shown in the Fig 1.

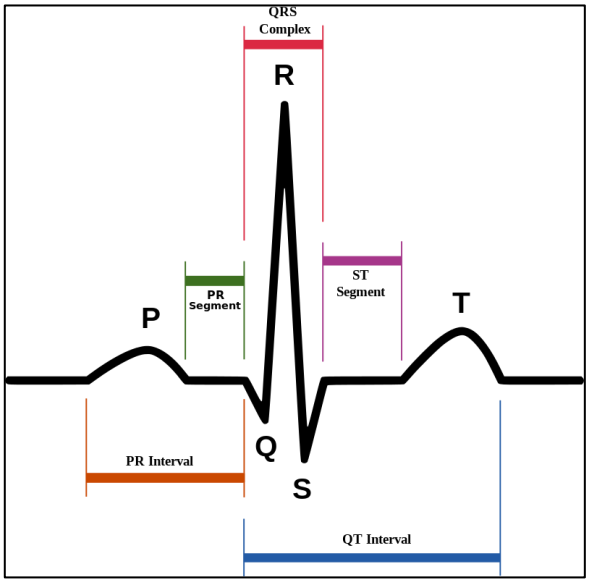


Figure 1. ECG of a heart in normal mode.

B. *Heart Rate Variation (HRV)*

The ECG Waveform varies according to a person's physical activities and is associated with heart rates. In driver drowsiness scenarios, the heart rate continuously alternates between sleep and awake states [14]; yet, by using ECG Waveform, the change in heart rate will be detected accurately. The heart rate variability is calculated by the variation in the time interval between consecutive R peaks of the QRS complex, and this variation is utilized to detect the driver's drowsiness. Heart Rate Variation (HRV) is shown in Figure 2.

C. *HRV Analysis*

Previous research confirms that the ratio of LF to HF is correlated to driver drowsiness. HRV analysis can be divided into two categories: time domain and frequency domain [15]. The time-domain approach is the most facile to perform considering that it is used on a series of consecutive RR intervals in order to obtain variables (e.g. the standard deviation of NN intervals (SDNN), the root mean square of consecutive differences (RMSSD), etc).

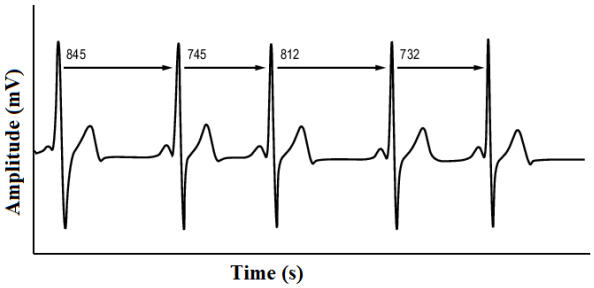


Figure 2. Heart Rate Variation.

In the frequency-domain approach [16], the RR interim series is transformed to an equally sampled series, and then, a Power Spectral Density (PSD) using Fast-Fourier Transform (FFT) is calculated. To further clarify, the PSD is split into 3 frequency bands: high frequency (HF) from 0.15 to 0.4 HZ, low frequency (LF) ranging 0.04 to 0.15 HZ, and the very low frequency (VLF) from 0.0033 to 0.04 HZ. For each frequency region, actual and relative powers are calculated and the ratio of low frequency (LF) to high frequency (HF) power is crucial for measuring parasympathetic activities.

D. *Noises in ECG signal*

Heart rate variation is highly prone to artifacts, and the smallest errors in 2% of the data generate undesirable biases in heart rate variation measurements; therefore, accurate measurements are made by reducing baseline, artifacts, electrode contact, and powerline noises before the heart rate variation analysis. Largely, these noises are either removed using the software or hardware filters.

1) Electrode Contact Noise and Muscle Artifacts: The Electrode contact noise is generated either due to the loss of contact between the electrode and the skin or the electrode's movement away from the contact area on the skin; contrarily, the artifact noise results from the electrical activity of the driver's muscle contractions. In truth, these noises result in abrupt changes in the ECG signal's amplitude thus producing errors. Nonetheless, these noises are approximately 30 HZ and are removed using a low pass filter realized in hardware with a specified cutoff frequency of 33HZ.

2) Power line noise: The noise generated from the power system is the major cause of noise in the process of monitoring ECG signals (also known as power line interference) [17]. It is 60HZ in the USA and 50HZ in other countries; furthermore, this noise is caused by the device's electromagnetic field, stray effects of ac current fields, electromagnetic interference of power line, and improper grounding of the ECG device or the subject. Thus, if this noise is not removed, it will corrupt ECG data.

4. Design And Implementation

A. Design of the system

A detailed explanation of the proposed architecture is shown in Figure 3.

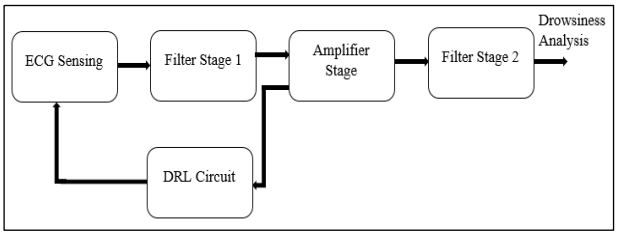


Figure 3. Drowsiness detection system architecture.

1) ECG Sensing: Traditionally, ECG devices have had electrodes, which are placed on the patient's skin in order to determine the heart's electrical activity, and recently, the use of ECG signals has extended into the technological field. Initially, analysis of ECG waveforms for driver drowsiness detection were performed using wet electrodes attached to the driver; however, in utilizing these electrodes for driver drowsiness detection, it has made it difficult to drive effectively. Therefore, it is crucial to eliminate the wires so that these drivers can drive unimpeded. Further research on the ECG compatible components required for drowsiness detection led to the development of dry or non contact sensors, which combats the complications that wet electrodes pose. In this paper, a PS25255 non contact sensor was used to capture the ECG required for drowsiness detection. These sensors are incorporated on to the seat back and seat belt with appropriate conductive fabric.

2) Filter Stage 1 : Filter stage 1 is employed at the input terminal after the ECG sensing stage in order to remove 30HZ of electrode contact noise and muscle artifacts. Due to the presence of these noises, the ECG signal is corrupted; hence, in order to remove the noise, a low pass filter with a cutoff frequency of 33HZ was used.

3) Amplifier Stage: In order to improve the strength of the ECG signal for accurate drowsiness detection results, we used the amplifier stage. In this paper, the amplifier stage was implemented in two stages. In the first stage of amplification, the input ECG signals are weak; therefore, it was observed that they should be strengthened in order to facilitate the analysis. In the second stage (also called the differential stage), the first stage's output is amplified by minimizing the sensor's common potential. Moreover, the process of diminishing the interference depends on the type of differential amplifier employed in the ECG detecting hardware's input stage. In this paper, we used the unity gain stable OPA 2277 for amplifier stage 1 because of its high common-mode rejection, output free from phase inversion, ease of use and high performance. OPA 177 for amplifier stage 2 because it is unity gain stable and has high performance.

4) Filter Stage 2: Filter stage 2 is employed after the amplifier stage to remove 60HZ of powerline noise, which results in false positives during the analysis. Thus, a twin-T notch filter with a 60HZ cut off frequency was used to eliminate the powerline noise; additionally, a variable quality factor is achieved with a potentiometer that not only enhances the efficiency of the filter but also reduces errors. Broadly, this adjustment will give flexibility to the board, and in our design, all the resistors contain 0.05% tolerance to avoid frequency drift from 60HZ. Here, we used OPA 2277 for filter stage 2 because it is unity gain, operates up to 36-V supply rails, ultralow offset voltage,offset voltage drift, and 1-MHz bandwidth.

5) Driven Right Leg (DRL) circuit: A Driven Right Leg Circuit (DRL) is added to the bio-amplifiers to decrease the common-mode interference. To clarify, Bio-amplifiers measure very low frequency signals produced by the body, and due to the electromagnetic interference, the body operates as an antenna and picks up the 60HZ power line noise. Furthermore, the DRL circuit has an ECG sensor that detects the ECG signals and eliminates interference.

B. Hardware Simulation

A drowsiness detection system is viewed as a fundamental system with inputs and outputs; thus, in order to run this system in a simulation environment, the environment itself must support inputs and outputs that are typically analogous to those of the real-time signals. Consequently, the motive is to simulate the complete hardware in a SPICE simulator of electronic circuits. In our research, we used LTSpice, a freeware computer software, which implemented a SPICE simulator. The step-by-step procedure for the simulation is shown in Figure 4.

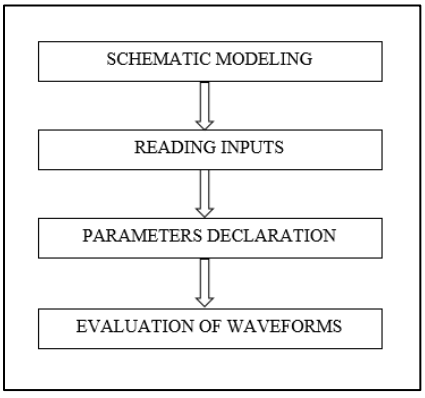


Figure 4. Simulation Steps

In LT Spice, the design model was initiated by a schematic capture. The input ECG samples were collected from different drivers from various driving scenarios, and the values that were obtained were converted into an excel file using MATLAB software. In the subsequent step, this file was converted into a text file and given as the input (it is necessary to note that the transient time and AC analysis are the parameters used for the simulation). The waveforms were generated for each of the design's stages, and they were then further analyzed for drowsiness detection.

1) Input Stage: In this simulation, the ECG signal with a peak amplitude of 150mV was applied differentially to the system's input; this signal comprises muscle noise and powerline noise. Figure 5 represents the input of the ECG signal applied to the circuit.

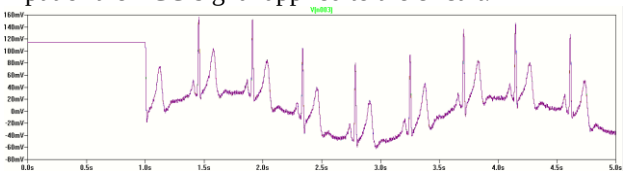


Figure 5. Input signal

2) Filter Stage 1: The applied input passes through two low pass filters with a cutoff frequency of 33HZ, and these filters suppress the interference due to muscle artifacts. In this stage, the peak output voltage of the ECG signal is 150mV. Figure 6 represents waveform at filter stage 1.

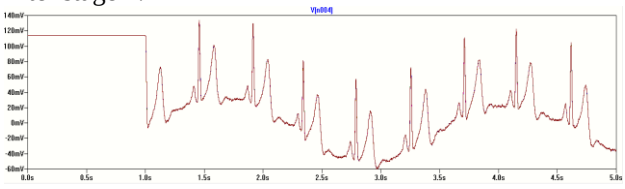


Figure 6. The filtered output from filter stage 1

3) Amplifier Stage 1: After passing through the low pass filter, the input then passes through amplifier stage 1. Principally, these amplifiers improve the strength of the two ECG signals, resulting in a peak output voltage of 1.4V. Figure 7 displays the output at amplifier stage 1.

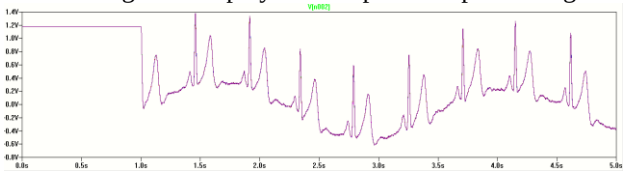


Figure 7. The output obtained from one of the stage 1 amplifiers

4) Amplifier Stage 2: Once the ECG signal passed through amplifier stage 1, the signal passes through a differential amplifier in order to minimize the common voltage between the two sensors. The peak output voltage of this stage is 2.8V. Figure 8 represents the output obtained at amplifier stage 2.

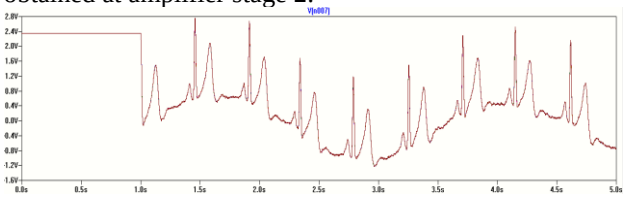


Figure 8. Output after amplifier stage 2.

5) Filter Stage 2: Due to the power line interference, ECG signals are mixed with a 60HZ power line noise; however, this noise can be removed using an active twin T notch filter with a cutoff frequency of 60HZ. After passing the output of amplifier stage 2 through a notch filter, powerline interferences are removed; hence, the peak output voltage of this stage is 2.8V. Figure 9 portrays the output after the notch filter stage.

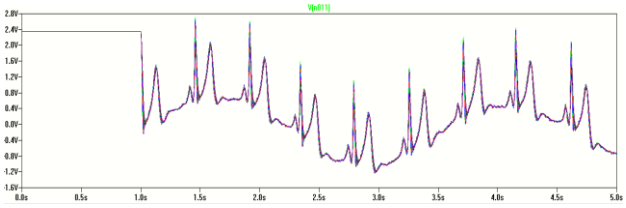


Figure 9. Output after notch filter stage.

6) Output: The output of the proposed circuit was adjusted by using a potentiometer at the output stage. The peak output voltage of this stage is 2.8V for diverse potentiometer values, and Figure 10 shows the potentiometer output.

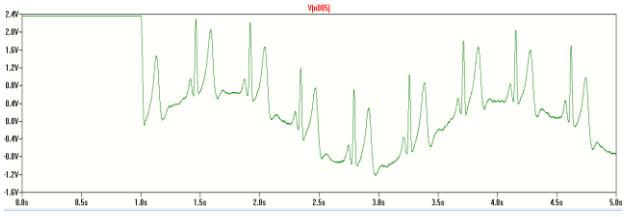


Figure 10. Output for a value of the potentiometer.

From Figure 10, it is evident that the proposed circuit produces an output ECG signal with peak amplitude of 2.8V, which is sufficient for drowsiness analysis.

Table 1 summarizes the output voltages obtained at each stage. From the table we can infer that the input was processed at each stage to obtain an amplified signal without noise. Additionally, the results suggest that the output produced from the proposed two-stage amplifier and filter is sufficient for detecting driver drowsiness.

TABLE I. OUTPUT VOLTAGES OBTAINED AT EACH STAGE

STAGES	OUTPUT VOLTAGE (V)
Input stage	0.150
Low-pass filter stage	0.150
Amplifier stage 1	1.4
Amplifier stage 2	2.8
Notch filter stage	2.8
Output	2.8

For the simulation, we collected data from different drivers for five-hour period including the subject's transition from awake state to asleep state and the data after sampling is then converted to discrete data. An algorithm for detecting HRV from the data and making the decision between the non-drowsy and drowsy states is done using logistic regression. The developed algorithm shows a consistent accuracy over 90% in 20 seconds. This algorithm was developed by our software team. [18]

5. CONCLUSION AND FUTURE WORK

Statistically, driver drowsiness is the leading cause of traffic accidents, and because of this, we need to employ a drowsiness measuring technique to subvert the prevalence of drowsy related driving accidents. In this paper, we detect drowsiness from ECG signals that are captured by non contact ECG sensors, and these collected ECG signals are amplified using two-stage operational amplifiers. The amplified signals are distorted by the presence of muscle artifacts and power line interference; thus, these noises are removed using filter stages 1 and 2 Within this research, a low-pass filter with a cutoff frequency of 33HZ was used to remove muscle artifacts, and a twin-T notch filter with a cutoff frequency of 60HZ was utilized to eliminate power line interference. The simulation was performed using LT Spice, and ECG signal graphs were plotted for the proposed design's multiple stages. The proposed methodology produces a maximum output of 2.8V, which proves adequate for drowsiness analysis. The results conclude that the proposed design is effective in generating the input signals required for drowsiness detection.

As a future work, the proposed design can be remodeled with different simulation tools, amplifiers, filters, and sensors to monitor driver drowsiness; moreover, the design's efficacy can be checked by placing the sensors in various positions. The proposed design can be implemented with a microcontroller along with an alarm and software algorithm in order to be used in real-time. In conclusion, the proposed method in this paper produced the maximum output voltage required for drowsiness detection and can be altered with new components for future developments.

ACKNOWLEDGMENT

The authors would like to thank Christopher Maye for his efforts to revise the paper and California State University Long Beach's Management for providing facilities to execute this work.

REFERENCES

[1] World Health Organization (WHO). (2015). Global Status Research on Road Safety 2015 [Online]. Available: http://www.who.int/violence_injury_prevention/road safety_status /2015/en/.

[2] J.Sztajzel, "Heart rate variability: a noninvaisive electrocardiographic method to measure the autonomic nervous system," Swiss Medical Weekly, vol. 134, pp.514-22, 2004.

[3] A. Muhamad, B. Nasreen, D. Michael, " A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability", in *Proc*. Sensors 2017.

[4] Sahayadhas. A, Sundaraj. K, and Murugappan. M, "detecting driver drowsiness based on sensors: a review", in *Proc.* Sensors 2012, pp-16937-16953

[5] K. Thomas, R. Andreas, and S. Nikoletta, "A robust drowsiness detection method based on vehicle and driver vital data", in Mensch und Computer 2017-Workshopband. Regensburg: 2017

[6] V. Swapnil Deshmukh, T.Bharti, A. Ansari, and M.S. Khatib, " Hardware implementation of driver safety system", in Proc. IOSR Journal of Computer Science, 2014, pp 79-82.

[7] Ghosh, Sayani, Nandy, Tanaya and Manna, Nilotpal. (2015). Real Time Eye Detection and Tracking Method for Driver Assistance System. [Online]. Available: 13-25. 10.1007/978-81-322-2256-9_2.

[8] Assari A. Mohammad, and Rahmati Mohammad, "Driver drowsiness detection using face expression recognition", in Proc. IEEE International Conference on Signal and Image Processing Applications, 2011.

[9] G. Rajendra Kumar, Dr. Samuel Vara Prasada Raju, and D. Santhosh Kumar, "Classification of ECG signals for drowsiness detection in brain and computer interface", in Proc. GESJ: Computer Science and Telecommunications, 2012.

[10] T.P. Deepa, and Reddy Vandana, " EEG based drowsiness detection using mobile device for intelligent vehicular system", in Proc. International Journal of Engineering Trends and Technology, 2013, Vol 6.

[11] Sang-Joong, Heung-Sub Shin, Wan-Young Chung "Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel" IET intell. Transp. Syst, 2014, Vol.8.

[12] X. Yun, "Real-time nonintrusive detection of driver drowsiness", unpublished.

[13] M. Sangeetha, S. Kalpanadevi, M. Rajendiran, and G. Malathi, " Embedded ECG based real time monitoring and control of driver drowsiness condition", in Proc. International Journal of Science, Technology and Society, 2015, pp-146-150.

[14] Kannan, G. Ramaswamy, A. Gujjar, and S. Bhaskar, "Drowsiness onset detection," ed: Google Patents, 2017.

[15] K. M. Lee, S. M. Lee, K. S. Sim, K. K. Kim, and K. S. Park, "Noise Reduction for Non-Contact Electrocardiogram Measurement in Daily Life," Computers in Cardiology, vol. 36, p. 493−496, 2009.

[16] P. T. Ahamed Seyd, V. I. Thajuddin Ahamed, J. Jacob, and P. K. Joseph, ”Time and frequency domain analysis of heart rate variability and their correlations in diabetes mellitus”, in Proc. World Academy of Science, Engineering and Technology, 2008.

[17] C. M. Gibbson, and C. Zorkun .(2012). EKG Interpretation Basics. [Online]. Available: http://www.wikidoc.org/index.php/EKG_ interpretation_basics.

[18] B. Mohsen, B. Nitish, E. Bianca and M. Mozumdar, “Real Time Driver Drowsiness Detection using a Logistic-Regression-Based Machine Learning Algorithm”, in *Proc.* IEEE Green Energy and Smart System Conference, 2016.

Srihitha Jujhavarapu is pursuing a Master’s degree from California State University Long Beach. She obtained a Bachelor’s degree in Technology from Jawaharlal Nehru Technological University India.

Mohsen Babaeian is pursuing a Ph.D from California State University Long Beach. He obtained a Master’s degree in Telecommunications from California state University Long Beach, and Claremont Graduate University. His interests include biomedical signal processing.

Dr. Mohammad Mozumdar received Ph.D. degree in electronics and communication engineering from the Politecnico di Torino, Italy. His novel ideas of model-based design for sensor networks made profound impact on engineering and industrial communities and have been published in book chapters, renowned journals, conference proceedings, major scientific magazines, and also translated in several different languages. He is a tenured Faculty with the Electrical Engineering Department, California State University at Long Beach, and an ex-post-doctoral fellow from the University of California, Berkeley. His research interests include methodologies and tools for embedded system design, in particular, in the domain of sensor networks; energy efficient building management and control system design; cloud computing; cyber physical system; and methodology for the design of distributed embedded systems subject to high real time, safety and reliability constraints.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.4 (July-2018)

http://dx.doi.org/10.12785/ijcds/070404

Developing RC4 Algorithm Using S-Box of Advanced Encryption Standard Cipher

Ali M. Sagheer [1], Sura M. Searan[1] and Salih S. Salih[1]

[1]*Department of Information Technology, University of Anbar, Anbar, Iraq*

Received 5 Dec. 2017, Revised 21 Jan. 2018, Accepted 15 Feb. 2018, Published 1 July 2018

Abstract: RC4 stream cipher is one of the most significant symmetric cryptosystems, it is simple and used in many commercial products. RC4 uses dynamic permutations and avoids using Linear Feed Back Shift Register (LFSR). It has many weaknesses, such as the tendency in the generated key stream that some key bytes are biased toward different values. This paper presents a new algorithm using S-box of Advanced Encryption Standard (AES) to solve the correlation between public known outputs of the internal state. The state table is filled from S-box values and additional swapping operations are used. The analysis of the proposed algorithm over variable key length produces key byte streams that have no single and double bias. This paper obtains a new algorithm that combines the efficiency of the RC4 and robustness of AES. The results show that the sequences that are generated by the developed RC4 are more random than the sequence that was generated by the RC4. Also, the developed algorithm demands little time more than RC4 execution time. Additionally, the developed algorithm is robust against most attacks, such as distinguishing attack and can be used in different protocols such as Secure Sockets Layer (SSL) Protocol, Oracle Secure SQL, and Wired Equivalent Privacy (WEP) Protocol.

Keywords: RC4, Stream Cipher, S-box, Key Scheduling Algorithm (KSA), Pseudo Random Generation Algorithm (PRGA), Advanced Encryption Standard (AES), Single Bias, Double Bias.

1. INTRODUCTION

Encryption is a process that transforms plaintext into cipher text. It is basically used to ensure confidentiality. Organizations and companies are encrypting their data before transmitting in order to ensure secure data transmission in a public channel. Cryptographic algorithms are designed to be characterized by high speed of implementation, lower size, less complexity, and larger degree of security. Conventional cryptographic algorithms are complex and take a higher amount of energy when they are used by resource constrained devices in order to provide secure communication. Indeed, public key algorithms are still not appropriate in tracer networks for many reasons, such as finite storage and higher usage of energy. Therefore, security systems should be based on a symmetric key cryptography, especially in the systems that have limited hardware resources [1]. The strength of a stream cipher is the random key stream that assures secure computation of the cipher. The cryptanalysis of stream ciphers is essentially focused on identifying non-random proceeding; till date, the analysis of stream ciphers has been employed to identify the happening of non-random proceedings [2]. The same algorithm is used for encryption and decryption; the plaintext stream is XOR-ed with the generated series of the random key generator. RC4 algorithm is used in many wireless network systems and protocols [3]. It is used in SSL protocol, Oracle Secure SQL, WEP Protocol; it is also used to protect wireless networks as part of WPA protocol and to protect the internet traffic as part of the TLS (Transport Layer Security) protocol [4]. There are many attacks presented to analysis RC4 by [5]. RC4 is analyzed by different cryptanalysis according to RC4 different weaknesses [6]. The modern researches proved that you can practically utilize single and double byte biases for RC4 to acquire any part of the Internet traffic, depending on TLS (Transport Layer Security) with RC4 option. The objective of this suggestion is to develop RC4 algorithm and analyzing the developed algorithm and shows that this algorithm is free from single and double bias while RC4 shows the bias that proved in the previous researches.

E-mail: ali.m.sagheer@gmail.com, surasms917@gmail.com, Salihsami90@gmail.com

2. Related Works

Many researchers work on analyzing the RC4 algorithm based on its weakness and suggest different algorithms. Prasithsangaree and Krishnamurthy (2003)[7] worked on analyzing RC4 and AES algorithms based on energy consumption. They determined that RC4 is more suitable for large packets, while AES symmetric algorithm is more suitable for small packets; further, RC4 is faster than AES. Maitra and Paul (2008) [8] worked on analyzing RC4 based on weakness in biases and proposed additional layers over the key scheduling algorithm and pseudorandom generation algorithm. In the same year, they determined the bias that can be perceived in S [S [y]] based on this form of permutation bias after the Key Scheduling Algorithm (KSA); a total work is presented to demonstrate that many key stream output bytes of RC4 are highly biased towards several linear collections of private key bytes (Maitra & Paul, 2008a) [5]. Al-Fardan et al. (2013) [2] determined the security of RC4 in Transport Layer Security (TLS) and Wi-Fi Protected Access (WPA) and applied single and double byte bias attack on RC4 and could retrieve some plain text bytes. In the same year, Hammood, Yoshigoe, and Sagheer (2013) [1] suggested RC4 stream cipher with two state tables (RC4-2S) as an enhancement for RC4. This enhancement solves the correlation problem between public known outputs of the internal state using permutation between (State1) and (State 2). In addition, the time period to generate the key of RC4-2S is faster than that original RC4, reduces the number of required operations in key generation. Also, Hammood, Yoshigoe, & Sagheer, 2015) [9] worked on enhancing security and speed of RC4 by proposing algorithms to enhance RC4, solve weak keys problems, and make it robust by using random initial state. The weaknesses in RC4 still represent an open challenge for developers.

3. Description of RC4 Cipher

The RC4 algorithm was proposed by Ron Rivest in 1987 and kept secret as a trade until it was leaked in 1994 [10]. It is a set of stream words of size *n*-bits [11]. RC4 starts with the permutation and uses a secret key to produce a random permutation with KSA. Based on the secret key, the next stage is Pseudo Random Generator Algorithm (PRGA) that generates key stream bytes which XORed with the original bytes of plaintext to produce the cipher text [8]. The state table is used to get pseudo-random bytes. This is done in the first phase of the algorithm [7]. The key is sometimes used as a 128-bit key. This operation is performed between key and plain text equivalent to Vernam cipher [12]. Many stream cipher algorithms use LFSR, especially in hardware architecture, but RC4 design does not. RC4 has a variable length of key that ranges between (0-255) bytes to initialize a 256-byte array in initial state (State [0] to State [255]) [1]. RC4 operated in two phases: the first consists in KSA, which initializes the internal state.

Algorithm 1. KSA of RC4
Input: Key
Output: State
1. For (i = 0 to 255) 1.1. State[i] = i 2. Set j = 0 3. For (i = 0 to 255) 3.1. j = (j + State[i] + Key [i mod key-length]) mod 256 3.2. Swap (State[i], State[j]) 4. Output: State

The second is a PRNG. It generates the output key stream.

Algorithm 2. PRGA of RC4
Inputs: State, Plaintext $_i$
Outputs: Key sequence (K sequence)
1. i = 0, j = 0 2. For (i = 0 to Plaintext length) 2.1. i = (i + 1) mod N 2.2. j = (j + State[i]) mod N 2.3. Swap (State[i], State[j]) 2.4. K sequence = State [State[i] + State[j]] mod N 3. Output: K sequence

The output sequence of key K is XORed with the plaintext

$C_i = K_i \oplus$ Plaintext $_i$ [13].

4. RC4 Weaknesses

The RC4 algorithm shows several weaknesses; some can be worked out, but others are difficult to resolve. One of these weaknesses in the initialization state is the statistical bias which occurs in distributing words of the first output. This bias makes it slight to distinguish between many short output of RC4 and random strings by analyzing their second word. This weakness is used to make effective cipher-text-only attack on this algorithm in broadcast applications, where the same plaintext is sent to multiple receivers with different keys. The unique statistical behavior is independent from the KSA and remains applicable even when the RC4 begins with a totally random permutation [14]. The slide in search effort from this attack is 25:1, but, when using linearly related session keys, the slide in effort increments to 218, that causes the weak keys [15]. Roos found weaknesses in RC4 that show a robust correlation between generated value and the first few values of the state table [16]. The main cause is the state table began in series (0, 1, 2, …., 255) and at least one out of every 256 possible keys, the

first byte of the generated key, is highly correlated with a few key bytes. So, the keys allow for the precursor of the first bytes from the PRGA output [9]. The goal of the attack is to retrieve the original key, the internal state, or the output key stream to have access to the original messages. From the previous studies based on KSA and PRGA, RC4 shows the following weaknesses: biased bytes, distinguishers, key collisions, and key recovery from the state [1]. Mantin and Shamir found the major weakness of the algorithm in the second round is the probability of zero output bytes [17]. Fluhrer and McGrew found a serious weakness: anyone who knows a portion of the private key can potentially attack fully on the RC4 [18]. Maitra and Paul found a secret key by using the initial state table. Specifically, these authors generated an equation on the basis of the initial state table, selected some bytes of the secret key based on their assumption, and found out the private key by using the equation [8]. The attack aims to retrieve the main key, the internal state, or the final key stream to access to the original messages [19].

5. THE PROPOSED ALGORITHM (RC4 WITH S-BOX OF AES)

This section presents a new development of the RC4 algorithm by using S-box of the AES algorithm. The idea of this proposition is taken from Rijndael algorithm. The substitution bytes of the AES is a nonlinear transformation that uses 16 bytes of S-Boxes tables, S-Box is the multiplicative inverse of a Galois field $GF(2^8)$ followed by affine transformation [7]. This suggestion aims to combine the robustness and the security of the AES algorithm with the speed and the simplicity of the implementation of the RC4. More in detail, the initial state table contents are substituted with the elements of S-box to eliminate the correlation between the internal state and public known output and to reduce the weakness that is exploited by the attacks by increasing the randomness and the complexity. This algorithm starts with the initialization KSA algorithm and then the PRGA algorithm, as shown in Figure 1 below. All operations are implemented mod State length. The KSA takes a secret key *k* with a 128 *n*-bit long word in the first step; the state tables are filled by numbers from 0 to N-1 and then substituted by S-box. The input secret key is used as a state table seed. After the KSA, the PRGA performs additional swapping operations between state[i] and state[i+1], and between state[j] and state[j+1], to generate the key stream that will XORed with the plaintext to get the cipher text.

The first phase is KSA:

Algorithm 3. KSA for RC4 with S-Box of AES

Input: Secret Key

Output: State

1. S-box [256] = S-box of AES algorithm
2. For (i = 0 to N – 1)

   2.1 State[i] = S-box(i)
3. Set j = 0
4. For (i = 0 to N – 1)

   4.1 j = (j + State[i] + Key [i mod key-length]) mod N

   4.2 Swap (State[i], State[j])
5. Output: State

The other is PRGA phase as shown below:

Algorithm 4. PRGA for RC4 with S-Box of AES

Inputs: State Table, Plaintext

Outputs: Key sequence (K), Ciphertext (C)

1. Initialization:
   - 1.1 i = 0
   - 1.2 j = 0
2. For (i = 0 to P_Length)
   - 2.1 i = (i + 1) mod N
   - 2.2 j = (j + S-box(j) + State[i]) mod N
   - 2.3 j2 = (j2 + S-box(j2) + State[i]) mod N
   - 2.4 Swap (State[i], State[j2])
   - 2.5 For (j = 0 to N – 1)
     - 2.5.1 Swap (State[j], State[j+1])
   - 2.6 K sequence = State [(State[i] + State[j] + S-box (j2 mod N)) mod N]
   - 2.7 $C_i = K_i \oplus P_i$
3. Output: K sequence and C_i

The model of double RC4 with S-box of AES is shown in figure 1.

Initial with numbers from 0 to State length. Fill with chosen key.

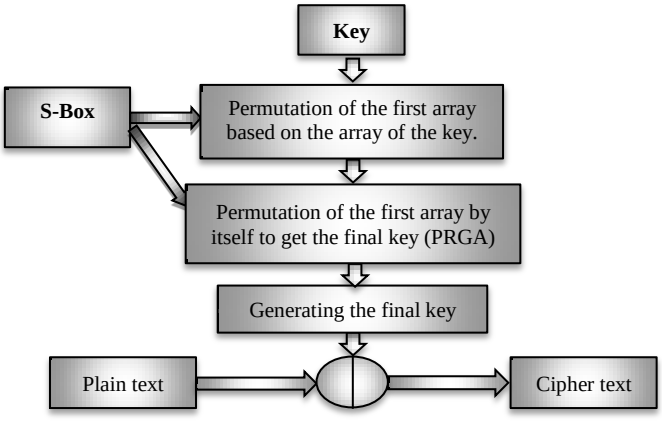


Figure 1. The model of developed RC4 encryption algorithm.

6. THE ANALYSIS OF RC4 AND DEVELOPED RC4 ALGORITHM BASED ON SINGLE BYTE BIAS

Mantin and Shamir (2002) [17] were the first researchers that denoted bias in the key stream of the RC4; their result was highly accurate. Sarkar, Gupta, Paul, and Maitra (2015) [4] determined a key-length-dependent bias in the key streams of the RC4 and worked with 256-byte keys. (AlFardan, Bernstein, Paterson, Poettering, & Schuldt, 2013) [2] denoted additional biases in the key stream of RC4 that do not have theoretical observations. In this work, these researchers analyzed the RC4 and the proposed algorithm. The proposed algorithm has no bias in the key distribution bytes as determined below as a result of the use of additional operations that cause no correlation between internal state and the output sequence. Algorithm 5 is used to measure the distribution of key stream bytes.

Algorithm 5. Measuring distributions of key stream bytes

Input: Key [k_1, k_2, …., k_{16}].

Output: Key position (Kp), Key value (Kv), and the number of Frequents in each position (Kf).

1. For (x = 1 to 2^{21}) Do
   - 1.1 i = 0
   - 1.2 j = 0
   - 1.3 Call Algorithm 1: KSA
   - 1.4 Call Algorithm 2: PRGA.
   - 1.5 Deducting new key with length = 16 from each generated key to be a new secret key.
2. For (i = 1 to N)
   - For (j = 1 to values. Count)
   - 2.1 If (values [i] == value)
   - 2.2 Increment count by 1
   - 2.3 Frequents = (count / (2^{21} * 16))
3. Return Kp, Kv, and Kf for each position of key stream bytes.

The state table is analyzed with 32 positions to reduce the search space and 2^{21} secret keys, each one with length 16, and produces 32 positions to calculate the frequent of each of the 32 values in each position. The key distribution bytes of the RC4 and the modified RC4 are determined in the following charts.

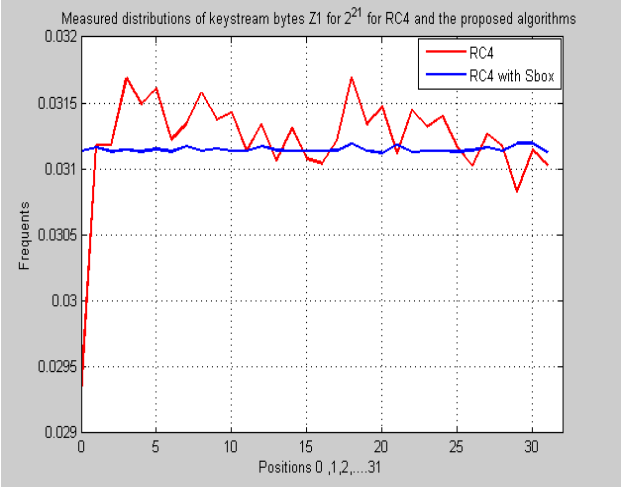


Figure 2. Key distribution bytes in the first position with 2^{21} for the RC4 and the developed RC4.

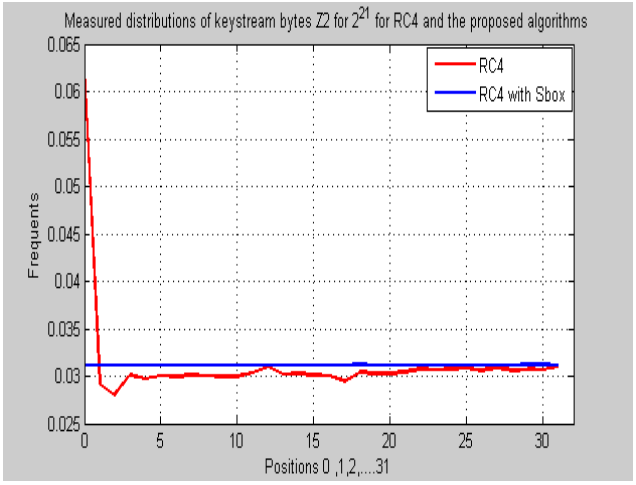


Figure 3. Key distribution bytes in the second position with 2^{21} for the RC4 and the developed RC4.

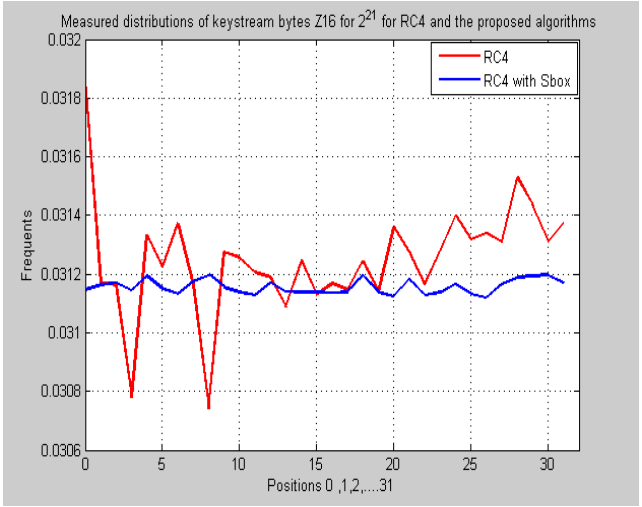


Figure 4. Key distribution bytes in the 16th position with 2^{21} for the RC4 and the developed RC4.

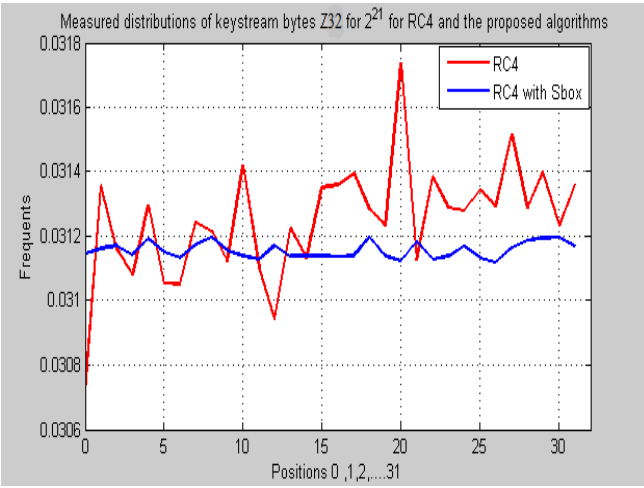


Figure 5. Key distribution bytes in the 32th position with 2^{21} for the RC4 and the developed RC4.

The distribution of key stream bytes of the RC4 algorithm has shown the same biases that are observed in literature. The experiment executed with generated key stream of 32 bytes and the number of generated keys ranging from 2^{16} to 2^{21} with independent random secret keys of 16 bytes. Expected biases started to appear for runtime beyond 2^{16} key generations, as shown in figure 2 to figure 5. They became apparent when the generated key stream increased to 2^{20}. The proposed algorithm shows that there is no bias in its key distribution bytes and the implementation time of the key stream generation is more than that required for the implementation of the RC4. The complexity and randomness in the proposed algorithm key is higher than the RC4 key bytes.

This algorithm is implemented in C#.Net programming language. Several biases were identified in literature. The RC4 successfully reproduced and proved these biases in the first 32 bytes of the key stream, while the developed RC4 has no bias in the first 32 positions of the key stream bytes.

7. The Analysis of RC4 and Developed RC4 Algorithm Based on Double Byte Bias

After explaining single-byte biases, that are significant to the cryptographic society, the attack simply can be avoided by ignoring the initial bytes. Thus, the RC4 with additional configuration can still be resistant to the single-byte bias attack. However, several authors have investigated biases beyond initial bytes and have discovered different multi-byte biases in the key stream of the RC4.

Fluhrer and McGrew (2001) [18] were the first researchers that discovered the biases in a consecutive pair of bytes (K_i, K_{i+1}) and detected long-term biases of the RC4. They discovered ten positive biases that mean their probability was higher than the desired value; besides, they detected two negative biases that mean their probability was lower than the desired value. Hammood and Yoshigoe (2016) [13] estimated the probability of the cipher for generating each pair of byte values through each 256 byte cycles and got a complete view of the distributions of every pair of byte values at the positions (i, i + 1). They replicated Fluhrer and McGrew's biases and indorse their work by Al-Fardan et al.'s (2013) studies. They found two new positive biases not mentioned by Fluhrer and McGrew (2001).

This work reproduced Fluhrer and McGrew's (2001) biases and Hammood and Yoshigoe (2016) bias with 1024 keys of 16 bytes to generate 2^{32} keystream bytes after discarding the first 1024 bytes. Each key from the 1024 keys generates 2^{32}; therefore, the whole amount of generated keys is 2^{42}. The developed RC4 using S-box did not generate any statistical bias and its output in the range only $\pm 2^4$ from the predicted occurrences. Algorithm 6 below determines the measuring of double byte bias. The main idea of this algorithm is to measure the appearance of the pair (Z_i, Z_{i+1}) in each position of the output of the RC4.

Algorithm 6. Measuring distributions of key stream bytes (K_a, K_{a+1})

Input: K [k_1, k_2, …., k_{16}]

Output: 3-Dimentions array

1. i = j = i1 = k = 0
2. For (x = 1 to 2^{10})
   - 2.1. Call Algorithm 2.1: KSA
   - 2.2. For (R= 1 to 2^{32})
     - 2.2.1. i = (i + 1) mod N
     - 2.2.2. j = (j + State[i]) mod N
     - 2.2.3. Swap (State[i], State[j])
     - 2.2.4. Generated Key = State[(State[i] + State[j]) mod N]
     - 2.2.5. A[k] [Generated Key] [i1] = A[k] [Generated Key] [i1] + 1
     - 2.2.6. Deducting new key with 16 bytes from each generated key to be a new secret key.
     - 2.2.7. k = Generated Key
     - 2.2.8. i1 = (i1 + 1) mod N
3. Return A[k] [Generated Key] [i1]

Figure 6 shows the distribution of (Z_r, Z_{r+1}) for all the first 32 bytes of RC4 where Z_r = i and Z_{r+1} = i to discover possible double-byte biases. Y-axis determines the frequents of each pair of values while the X-axis contains each pair of values.

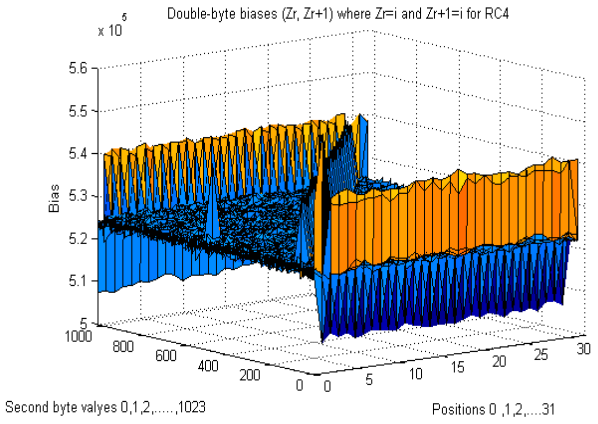


Figure 6. Double-byte biases (Z_r, Z_{r+1}) for RC4 where Z_r=i and Z_{r+1}=i.

Figure 7 shows the distribution of (Z_r, Z_{r+1}) for all the first 32 bytes of the developed RC4 with AES S-box where Z_r = i and Z_{r+1} = i.

Y-axis determines the frequents of each pair of values while the X-axis contains each pair of values.

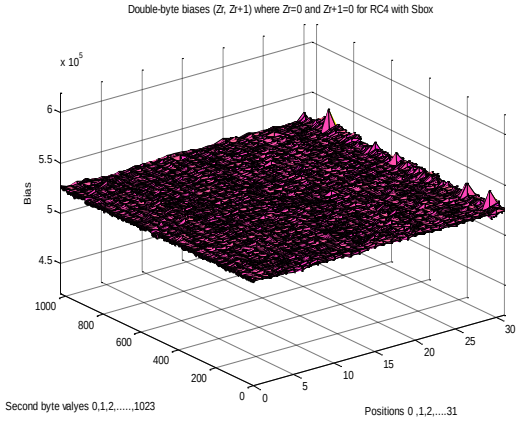


Figure 7. Double-byte biases (Z_r, Z_{r+1}) for RC4 with AES S-box where Z_r=i and Z_{r+1}=i.

8. RANDOMNESS TEST

The key stream generated by the RC4 and the developed RC4 was tested by the NIST (National Institute of Standards and Technology) Test Suite. The NIST is a statistical group for random number generator tests that consist of sixteen statistical tests to measure the randomness of output sequences of pseudo-random number generators or true random number generators, as shown below. The tests of this PRNG were done by using NIST STS-1.6.

The good random number generator likelihood was represented by the P-value in the test; this P-value was compared to 0.01. If the value is higher than 0.01, then the series is accepted, otherwise it is rejected because it shows no randomness. Some tests accepted large series sizes and failed in small series sizes, while other tests accepted both. In this paper, a large size (12 kilobyte) is generated from each secret key that has been used. These series are tested and the average of p-values results are calculated from these tests. As table I shows, the p-values are succeeded and the obtained series are uniformly distributed and random. If the tests give p-value equal to 1, then the sequence has complete randomness for this test. A p-value of zero means that the sequence has fully nonrandom.

TABLE I. RESULTS OF RUNNING NIST ON THE GENERATED KEY BY RC4 AND RC4 WITH AES S-BOX.

Test No.	**Statistical Test Name**	**RC4**		**RC4 with AES S-box**	
		P-VALUE	***Conclusion***	***P-VALUE***	***Conclusion***
1	Approximate Entropy	0.805578	SUCCESS	0.687713	SUCCESS
2	Block Frequency	0.742455	SUCCESS	0.621580	SUCCESS
3	Cumulative Sum(Forward)	0.739164	SUCCESS	0.464227	SUCCESS
4	Cumulative Sum (Reverse)	0.854066	SUCCESS	0.311231	SUCCESS
5	FFT	0.279715	SUCCESS	0.913344	SUCCESS
6	Frequency	0.898580	SUCCESS	0.481208	SUCCESS
7	Lempel-Ziv compression	0.889521	SUCCESS	0.453945	SUCCESS
8	Linear Complexity	0.407918	SUCCESS	0.842261	SUCCESS
9	Longest Runs	0.767817	SUCCESS	0.913467	SUCCESS
10	Non periodic Templates	0.540708	SUCCESS	0.570862	SUCCESS
11	Overlapping Template	0.497550	SUCCESS	0.597580	SUCCESS
12	Random Excursions	0.528198	SUCCESS	0.402825	SUCCESS
13	Random Excursion Variant	0.525591	SUCCESS	0.497233	SUCCESS
14	Rank	0.610871	SUCCESS	0.321188	SUCCESS
15	Runs	0.115965	SUCCESS	0.903451	SUCCESS
16	Serial	0.646168	SUCCESS	0.763967	SUCCESS
17	Universal Statistical	0.380374	SUCCESS	0.074774	SUCCESS

"Success" indicates that the series is acceptable and has good randomness, while "Failure" indicates that the series is not acceptable and not random.

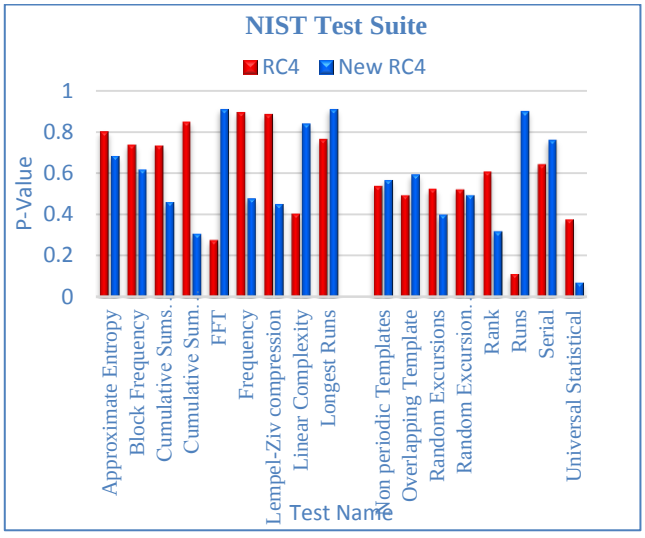


Figure 8. Results of running the NIST suite on the generated key by the RC4 and the proposed RC4.

When the implementation of the proposed algorithm takes place on the same size of the secret keys, the developed algorithm is faster than the implementation of the AES and it requires less time than that required for the RC4, as Table II shows.

TABLE II. KEY GENERATION TIME FOR THE RC4 AND THE PROPOSED RC4

Key Size	RC4 Time (m.s.)	RC4 with AES S-Box Time
301 bytes	1004	1008
3 KB	1033	1051
300 KB	35557	37223
3 MB	3202309	3421215

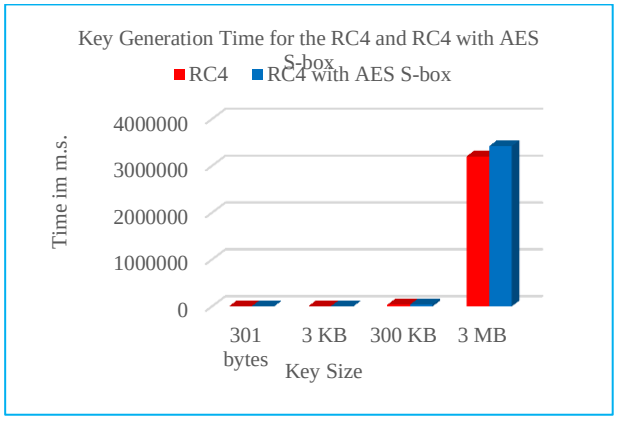


Figure 9. Key generation time for RC4 and RC4 with AES S-box.

9. CONCLUSIONS

In this paper, a new algorithm is proposed as a development for RC4 algorithm. RC4 is one of the most important symmetric cryptographic algorithms. The key generation phases are weak in key stream distribution bytes that biased toward different values. The proposed algorithm used the S-box of the AES algorithm to combine the speed of the RC4 algorithm and the robustness of the AES algorithm. The analysis of the RC4 and of the developed RC4 algorithm highlighted that the new algorithm has no single and double bias in the key stream. The developed algorithm requires little time more than that required for the RC4 by using additional swapping operations, but it is faster than the AES. The developed algorithm has passed all the statistical tests in the NIST suite, and can be used in different protocols such as SSL, WEP, and WPA protocol. As a future work, parallel processors may be used for implementing the analysis of RC4 and the developed algorithm with more key stream bytes (256 and more), and may compare the proposed scheme with other algorithms (such as DES, 3DES).

REFERENCES

[1] M. M. Hammood, K. Yoshigoe, and A. M. Sagheer, "RC4-2S: RC4 Stream Cipher with Two State Tables," Information Technology Convergence, Lecture Notes in Electrical Engineering, Springer Science Business Media Dordrecht1, pp. 13-20, DOI: 10.1007/978-94-007-6996-0_2, 2013.

[2] N. J. Al-Fardan, D. J. Bernstein, K. G. Paterson, B. Poettering, and J. C. Schuldt, "On the Security of RC4 in TLS and WPA," In Presented as Part of the 22nd USENIX Security Symposium,13, pp. 305-320, 2013.

[3] M. McKague, "Design and Analysis of RC4-Like Stream Ciphers," (Master Thesis), University of Waterloo, Canada, Ontario, 2005.

[4] S. Sarkar, S. S. Gupta, G. Paul, & S. Maitra, "Proving TLS-attack related open biases of RC4," Designs, Codes and Cryptography, vol. 77, no. 1, pp. 231-253, 2015.

[5] S. Maitra, & G. Paul, "New form of permutation bias and secret key leakage in keystream bytes of RC4," In Fast Software Encryption, vol. 5086, pp. 253-269, 2008. Springer Berlin Heidelberg.

[6] L. L. Khine, A New Variant of RC4 Stream Cipher. Mandalay Technological University Mandalay 05052, Mandalay, Myanmar: World Academy of Science, Engineering and Technology, 2009.

[7] P. Prasithsangaree, and P. Krishnamurthy, "Analysis of Energy Consumption of RC4 and AES Algorithms in Wireless LANs," In Proceedings of Global Telecommunications Conference, IEEE, vol. 1443, no. 3, pp. 1445-1449, December 2003.

[8] S. Maitra, and G. Paul, "Analysis of RC4 and Proposal of Additional Layers for Better Security Margin," Lecture Notes in Computer Science, International Conference on Cryptology, vol. 5365, pp. 27-39, 2008.

[9] M. M. Hammood, K. Yoshigoe, and A. M. Sagheer, "RC4 stream cipher with a random initial state," Proceedings in 10th FTRA International Conference on Secure and Trust Computing, data management, and Applications, Lecture Notes in Electrical Engineering, pp. 407-415, 2013, Springer Netherlands.

[10] S. Paul, and B. Preneel, "Analysis of Non Fortuitous Predictive States of the RC4 Keystream Generator," Springer Computer Science, vol. 2904, pp. 52-67, 2003.

[11] M. A. Orumiehchiha, J. Pieprzyk, E. Shakour, and R. Steinfeld, "Cryptanalysis of RC4 (n, m) Stream Cipher," In Proceedings of the 6th International Conference on Security of Information and Networks, vol. 178, pp. 165-172, 2013.

[12] P. Sepehrdad, S. Vaudenay, & M. Vuagnoux, "Discovery and Exploitation of New Biases in RC4," In Selected Areas in Cryptography, vol. 6544, pp. 74-91, 2011. Springer Berlin /Heidelberg.

[13] M. M. Hammood, and K. Yoshigoe, "Previously overlooked bias signatures for RC4," In Proceedings of International Symposium on Digital Forensic and Security, pp. 101-106, 2016. doi: 10.1109 / ISDFS.2016.7473526. IEEE.

[14] M. Robshaw, and O. Billet, "New Stream Cipher Designs: The eSTREAM Finalists," Lecture Notes in Computer Science, Springer Berlin/Heidelberg, vol. 4986, pp. 244–266, 2008, Springer, Heidelberg.

[15] S. Mister, and S. Tavares, "Cryptanalysis of RC4-like Ciphers," In Selected Areas in Cryptography, vol. 1556, pp. 131-143, 1999, Springer Berlin/Heidelberg.

[16] A. Roos, "A class of weak keys in the RC4 stream cipher," South African, Vironix Software Laboratories: Westville, 1995. Available at http:// http://marcel.wanda.ch/Archive/WeakKeys.

[17] I. Mantin and A. Shamir, "Practical Attack on Broadcast RC4," In Fast Software Encryption, Lecture Notes in Computer Science, Springer, vol. 2355, pp. 152-164, 2002.

[18] S. R. Fluhrer, and D. A. McGrew, "Statistical Analysis of the Alleged RC4 Keystream Generator," Lecture Notes in Computer Science, Springer- Berlin Heidelberg, vol. 1978, pp. 19-30, 2001.

[19] S. M. Searan, A. M. Sagheer, and M. M. Hammood, "Analyzing of RC4 Algorithm Based on Its Single and Double Byte Bias by Using New Algorithms," International Conference on Change, Innovation, Informatics and Disruptive Technology, London – UK, 11-12 OCT, 2016, Available at http://sriweb.org/londonconf/.

Ali M. Sagheer is a Professor in the Computer College at Al-Anbar University. He received his B.Sc. in Information System (2001), M.Sc. in Data Security (2004), and his Ph.D. in Computer Science (2007) from the University of Technology, Baghdad, Iraq. He is interested in the following fields; Cryptology, Information Security, Number Theory, Multimedia Compression, Image Processing, Coding Systems, and Artificial Intelligence. He has published many papers in different scientific journals.

Sura M. Searan has received her B.Sc. in Computer Science (2013) and M.Sc. in Information Security (2016) from the College of Computer Sciences and Information Technology at University of Anbar, Baghdad, Iraq. She is interested in the following fields: Cryptology, Information Security, and Coding Systems.

Salih S. Salih has received his B.Sc. in Computer Science (2012) and M.Sc. in Computer Science (2016) from the College of Computer Sciences and Information Technology at University of Anbar, Baghdad, Iraq. He is interested in the following fields: Coding Systems, Database, and Data Mining.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.4 (July-2018)
http://dx.doi.org/10.12785/ijcds/070405

IEEE 802.11 DCF Improvement: Waiting DIFS while Waiting Back-off

Latifa Souad Mahi-Rekik and Malika Bourenane[1]

[1]*University of Oran1 Ahmed Ben Bella, Oran, Algeria*

Received 28 Jan.2018, Revised 10 Mar. 2018, Accepted 11 Jun. 2018, Published 1 July 2018

Abstract: With the emergence of time-sensitive applications such as games and telephony, the introduction of the Quality of Service (QoS) in general and the improvement of transmission delays in particular have become a must in wireless networks. The distributed coordination function (DCF) being the fundamental access method and the basis of the wireless LANs IEEE 802.11 MAC protocol, a great number of works have been done to improve it. In DCF, after sensing an idle channel, stations have to wait before the transmission of each frame a length of time called DIFS (DCF Inter Frame Space) followed by another called back-off. If the medium becomes busy during the back-off process, the back-off timer is paused and resumed when the medium is sensed free for a DIFS again. This time loss becomes considerable when the number of interruptions of back-off process grows. This paper proposes a contribution to improve DCF by combining IFS and back-off time. The simulation results show that the approach benefits are proportional to the contention level of the network and to the number of hops in multi-hop network topologies.

Keywords: IEEE 802.11, CSMA/CA, DCF, IFS, DIFS, Back-off algorithm, contention window

1. INTRODUCTION

The fundamental access method of the IEEE 802.11 MAC protocol is a distributed coordination function (DCF) also called CSMA/CA for Carrier Sense Multiple Access with Collision Avoidance. The distributed CSMA/CA algorithm requires a gap of a minimum specified duration (called Inter Frame Spacing-IFS) between the different frames transmitted in order to establish a priority system between frames. Control frames such as acknowledgment (ACK) or clear to send (CTS) for example are given priority higher than that of data frames by waiting smaller IFS. The sending station checks if the channel has remained free during this time before it can transmit its frame. If the channel is busy, the station must delay its transmission by choosing a random number called back-off in an interval called contention window; this will determine an additional waiting time in order to solve partially channel access conflicts.

The objective of this work is to minimize wait times which lead to under usage of the channel by combining IFS and back-off times. In our approach, the DIFS time is eliminated whenever the back-off time is greater than or equal to DIFS (DCF IFS for Data frames); thus, the station is not obliged to wait for the channel to be free for DIFS time since its back-off time already includes it. However, DIFS wait is kept in case of back-offs that are smaller than DIFS. The approach is simulated using NS2 simulator and tested on different contention level scenarios of ad hoc single-hop and multi-hop 802.11 networks. Compared to the standard DCF, better performance is noticed in terms of throughput and end-to-end delay.

The rest of this paper is organized as follows: section II gives a brief description of DCF mechanism focusing on its inter-frame spacing and back-off mechanisms. Section III presents some researches on DCF and back-off algorithms improvement. Details and explanations of our contribution are given in section IV. Section V is devoted to implementation on NS2 simulator. Section VI presents simulation results and their interpretation for different topologies and scenarios. Section VII concludes the paper.

2. IEEE 802.11 MAC SUBLAYER

MAC sublayer of stations operating in an IEEE 802.11 LAN proposes three coordination functions which control access to the wireless medium: (1) DCF, the standard basis, (2) Hybrid Coordination Function (HCF) present only in QoS stations and (3) Point Coordination Function (PCF) optional, used for contention-free services. The IEEE 802.11 MAC architecture is described in Fig.1 as providing the PCF and HCF through the services of DCF [1].

The DCF coordinates the access to a shared medium (more exactly channel) by multiple stations. DCF is a CSMA/CA access mechanism. Like Ethernet, the station first checks that the channel is clear before transmitting. DCF defines two access mechanisms for packet transmission: (1) the basic one called the two way handshaking technique where the sender transmits data and the receiver responds with an ACK; and (2) the optional one called RTS/CTS technique or the four handshaking technique where the sender first transmits a short Request to send frame (RTS) and waits for a Clear-

E-mail:mahi.latifa@univ-oran.dz, mb_regina@yahoo.fr

To-Send frame (CTS) from the destination before transmitting data and receiving ACK.

This last technique is used to prevent hidden node problems but causes additional transmission delays. The rest of the paper focuses only on the basic technique.

DCF is based on a two-type time delay principle : (1) the inter frame spacing (IFS) to establish a priority system between frames of different natures,and (2) the random back-off timing to establish a priority system between stations which want an access to the channel simultaneously.

A. Inter frame spacing

The time interval between frames called the IFS plays an important role in coordinating access to the transmission medium. DCF uses five different inter frame spaces (see tables I and II). Varying inter frame spaces creates different priority levels for different types of traffic. The logic behind this is simple: high-priority traffic doesn't have to wait that long once the medium becomes idle. Therefore, if there is any high-priority traffic waiting, it grabs the network before low-priority frames have a chance to try [1, 2].

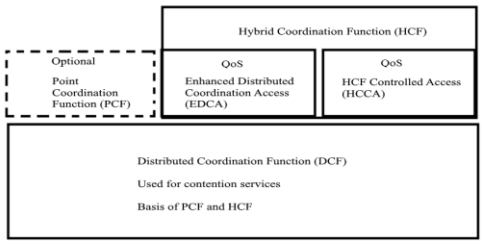


Figure 1 IEEE 802.11 MAC architecture

TABLE I DIFFERENT IFS AND THEIR USE

IFS	Name	Used by
SIFS	Short Inter frame Space	CTS, ACK, Fragments of a frame
PIFS	PCF Inter frame space	Point coordinated traffic
DIFS	DCF inter frame space	Data and management frames after a correctly received frame
AIFS	Arbitration inter frame space	QOS stations
EIFS	Extended Inter frame space	After an incorrectly received frame

TABLE II SOME IFS VALUES

		PHY Value Examples (in µs)		
Timings	**Value**	**FHSS**	**DSSS**	**OFDM**
Slot time	PHY dependent [1]	50	20	9
SIFS	PHY dependent [2]	28	10	16
PIFS	SIFS+ SlotTime	78	30	25
DIFS	SIFS+2*SlotTime	128	50	34

1. A Slot Time=Minimum time required for the PHY to determine the state of the channel +time to turnaround from receive to transmit mode +air propagation time +MAC processing delay.

2. SIFS=Time required to pass channel information between PHY sub-layers and between PLCP and MAC +time to turnaround from receive to transmit mode +MAC processing delay.

B. Backoff timing

The IEEE 802.11 Standard defines the Binary Exponential Back-off (BEB) algorithm to be performed in the following cases: (1) when the station listens to the medium before the first transmission of a packet and the medium is busy. (2) After each retransmission. (3) After a successful transmission. Whenever a data frame is to be sent, the station senses the medium; if it is free for at least a DCF inter-frame space (DIFS) period of time, the back-off mechanism is not used and the frame is directly transmitted. Otherwise, if the medium is busy, a back-off time B (measured in time slots which depend on the characteristic of physical layer) is chosen randomly in the interval [0, CW], where CW is called the contention window. After the medium has been detected idle for at least a DIFS, the back off timer is decremented by one for each time slot the medium remains idle [3].

If the medium becomes busy during the back off process, the back off timer is paused, and is resumed when the medium is sensed free for a DIFS again. When the back-off timer reaches zero, the frame is transmitted. Fig. 2 below illustrates the back-off process of two stations wanting to reach the channel simultaneously.

On the first transmission attempt, CW is set to a minimum value CWmin and at the next times (at the event of a collision), CW is doubled until it reaches a maximum value CWmax i.e. CW=min (2*CW, CWmax) . A new back-off time is then chosen and the back-off procedure starts over. After a successful transmission, the contention window is reset to CWmax.

We can design the Back-off algorithm as presented in Algorithm I below.

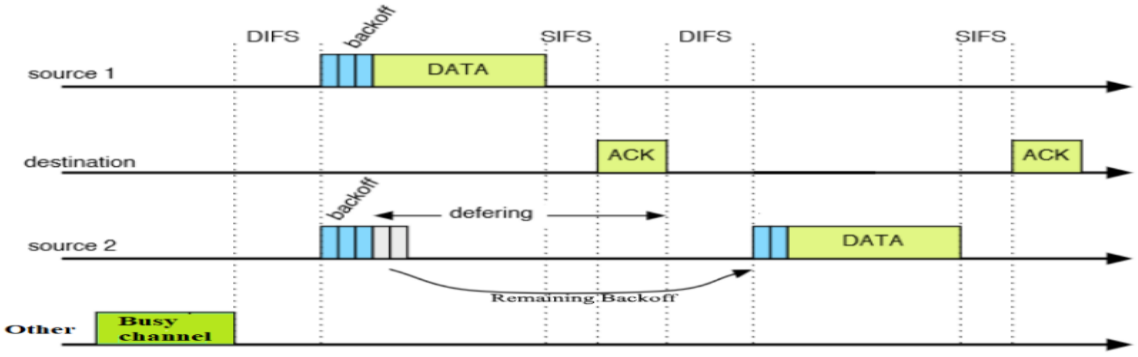


Figure 2 Back-off mechanism [4]

Algorithm I Back-off Algorithm in DCF transmission cycle

If (first transmission) and (channel free for DIFS) go to 4
CW=CWmin
For each packet transmission or retransmission
1. Wait for channel to be free for a DIFS
// DIFS wait before back-off start
2. N← rand(0,CW)
// Back-off time initialization (in time slots)
3. Do
   - Wait a slot_time
   - If (channel has been free during the whole slot_time)
     - N←N-1
   - Else
     // Channel busy, Back-off interrupted (frozen)
     - Wait for channel to be free for a DIFS
       // DIFS wait before back-off resume

While (N>0)
4. Transmit.
5. If (successful transmission) CW←CWmin;
else CW←min(2*CW, CWmax)

3. Related work

DCF being the fundamental access method and the basis of the IEEE 802.11 MAC protocol, many works have shown great interest to improve it. The reminder of this section classifies the contributions into two main categories:

A. Introduction of a certain quality of service :

DCF is unable to provide the required performance for voice and video applications, because it is mainly developed for best effort services. Basically, service differentiation at the MAC level is achieved by two main methods: priority and fair scheduling. The former binds channel access to different traffic classes by prioritized contention parameters; the latter partitions the channel bandwidth fairly by regulating wait times of traffic classes in proportion according to given weights. The tunable parameters for both approaches are CW size, back-off algorithm, and inter-frame space[5]. Some specific service differentiation mechanisms are:

1) *Enhanced DCF (EDCF*):

Part of IEEE 802.11e, EDCF prioritizes traffic by including Arbitrary IFS (AIFS) and minimum and maximum back-off window sizes; in EDCF, traffics keep using the same DCF access mechanism but have different probabilities of winning the channel. The IEEE802.11e amendment was approved in order to provide QoSsupport to WLANs. It defined the Hybrid Coordination Function (HCF) as an enhanced medium access mechanism which includes two access mechanisms that are: Enhanced Distributed Coordination Access (EDCA) and HCF Controlled Channel Access (HCCA). Although this amendment introduces the service differentiation scheme, it was not able to guarantee QoS for applications having strict QoS requirements [6].

2) Distributed Fair Scheduling (DFS) and Optimal DCF (O-DCF):

The main idea of DFS [7] is to differentiate the back-off interval (BI) based on the packet length and traffic class. In DFS, the station with smaller BI transmits first. O-DCF[8]controls link access aggressiveness by both CW size and transmission length based on the state of the queue. In O-DCF, links with a greater queue length are prioritized by decreasing their CW size and/or by increasing their transmission length.

B. DCF enhancement by changing backoff algorithm:

BEB is the key component of the DCF mechanism; however, it suffers from certain problems including significant delay degradation in case of saturated networks. Several proposals of back-off schemes have been made in order to solve the problem of exponential increase of the contention window after each failed transmission (generating useless access delays) and thus, providing better delay performance; we can cite for example:

1) *MILD (Multiplicative Increase Linear Decrease):*

MILD is a back-off algorithm where the multiplicative factor is 1.5 (instead of 2 in BEB); in MILD, the back-off upper bound (CW) is set as follows:

CW=min(1,5*CW,CWmax) after each failed transmission [9,10].

2) PBA (Padovan Backoff Algorithm):

In PBA which is based on the Padovan sequence, CW takes, after each failed transmission, the next Padovan term (CW=min(P(r),CWmax)), where r is the retry count and P(r) is the Padovan term [11].

In their study of back-off design for IEEE 802.11 DCF , Xinghua and Lin categorize back-off schemes into two groups (aggressive back-off with $\lim_{i\to\infty}(CW_{i+1}/CW_i)>1$and mild back-off with $\text{Lim}_{i\to\infty}(CW_{i+1}/CW_i)=1$). They show that aggressive back-off schemes such as BEB suffer from delay degradation when the network size is large.

Our contribution attempts to minimize 802.11 DCF channel access delay by combining inter frame space time and back-off time. In what follows, our approach is named DIB_DCF for DIFS In Back-off DCF and the original IEEE 802.11 DCF ORG_DCF.

4. PROPOSED APPROACH

As explained in section 2.B (back-off timing), a station wanting to transmit must defer its transmission for an additional time equal at least to DIFS (wait until the channel remains free for at least DIFS) before beginning its back-off process.Once the back-off process begins, it is decremented by one slot as long as the channel remains free; if it becomes busy again, the back-off process is frozen and resumes after a period of idle channel equal at least to DIFS and the operation is repeated until the back-off reaches zero.

Suppose a back-off process which has been interrupted n times before it reaches zero (Fig. 4 a), theoretically, the minimum time waited by the station (wait while channel is free) in this case is:

minwait_ORG=DIFS+BO1+DIFS+BO2+DIFS+…+DIFS +BOn

minwait_ORG= n*DIFS +$\sum_{i=1}^{n} BO_i$ =>

minwait_ORG=n*DIFS+BO (1)

Where BOi is the back-off portion elapsed before the interruption i.

And $\sum_{i=1}^{n} BO_i$=BO (Back-off time).

Our approach proposes to:

- Eliminate the DIFS wait before the first packet transmission and before each packet retransmission (back-off start) whenever, the initial back-off time (BO) is greater than or equal to DIFS.
- Eliminate the DIFS wait before each back-off resume whenever the remaining back-off time(RB) is greater than or equal to DIFS.

The minimum wait becomes in this case:

If (RB>DIFS) min_wait_DIB=$\sum_{i=1}^{n} BO_i$

Else min_wait_DIB=($\sum_{i=1}^{n} BO_i$)+DIFS

In other terms :

If (RB>DIFS)min_wait_DIB= BO time

else min_wait_DIB=BO time +DIFS (2)

From (2) and (1), we deduce that the gain is at least equal to (n-1)*DIFS.

Fig. 3 illustrates the original DCF and the new DCF mechanisms

Fig. 4 b illustrates the new DCF transmission cycle.

Algorithm 2 gives the new steps for DCF transmission cycle.

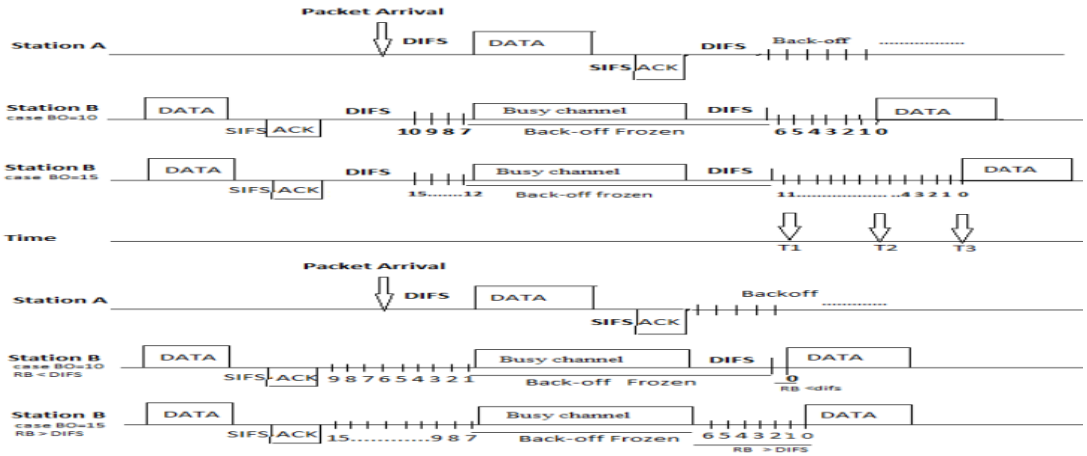


Figure 3 Illustration of ORG_DCF and DIB_DCF

The upper part of Fig. 3 shows the 802.11 DCF process in an example of two stations A and B transmitting in a wireless channel; the lower part shows our approach with the same example. Between both parts, we can see a time axis where the beginning times of DATA transmissions are reported.We can notice that, in both cases of station B (Remaining Back-off (RB)< DIFS and RB > DIFS), the station reaches the channel in a shorter time; it transmits data at time T1 in case of RB< DIFS and at T1+one slot time in case of RB > DIFS while with the original DCF, station B transmits respectively at T2 and T3.

Algorithm II DIB-DCF cycle Algorithm

For each packet transmission or retransmission

1. N← rand(0,CW)
   // Back-off time initialization (in time slots)
2. Rtime← N*slot_time
   //Back-off time in µs)
3. If (Rtime<DIFS)
   - Wait for channel to be free for a DIFS
     // DIFS wait before Back-off start only in case of BO<DIFS
4. Do
   - Wait a slot_time
   - If (channel has been free during the whole slot_time)
     - N←N-1
   - Else
     // Channel busy, Back-off interrupted (frozen)
     - If (Rtime<DIFS) Wait for channel to be free for a DIFS
       // DIFS wait before back-off resume only if RB<DIFS

   While (N>0)
5. Transmit

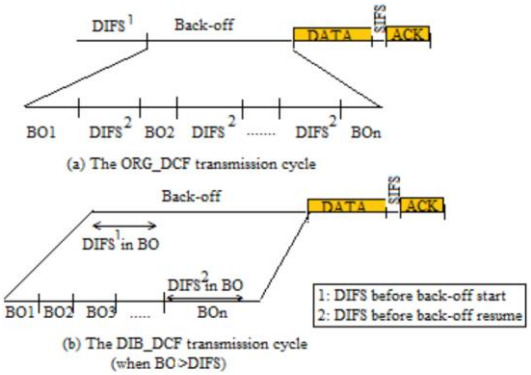


Figure 4 Transmission cycles in ORG and DIB DCF

5. SIMULATION OF THE PROPOSED APPROACH

In this section, NS2 simulator is described and the main changes made in its code to implement our proposal are presented. Simulation parameters are also given.

NS2 simulator is a popular discrete event network simulator developed under several previous research grants and activities; it remains in active use and will continue to be maintained [15].

A. Basic principles of NS2

NS2 is a C++ executable program which we call with a TCL configuration file as a parameter; Fig.5 describes the basic architecture of NS2. A TCL configuration file (called TCLsimulation script) is written in the interpreted Tool Command Language-TCL to describe the network to simulate (number of stations called nodes in NS2, type of links between nodes -duplex link, half duplex,…-, applications attached to nodes -constant bit rate, variable bit rate…- , type of physical layer, MAC layer, time at which transmissions begin, etc.). One simulation script describes one network scenario.

The C++ code (object oriented) contains the modelization of different components and protocols implied in the OSI layers of wired and wireless networks (propagation model, MAC layer 802.11, application layer, etc.) in addition to special components like timers or random number generators.

From the initial scenario, NS2, a discrete event simulator, creates a list of events with their execution times (the execution of an event consists in executing its associated actions). The simulation process then consists in executing the events in ascending order of execution times. In order to delay events (delaying events consists in scheduling them for a specified time), NS2 simulator uses the special components called timers. Class back-off timer is used to implement back-off timing and class defer timer to implement inter frame spacing delay. For a more complete and detailed presentation of NS2 see [13].

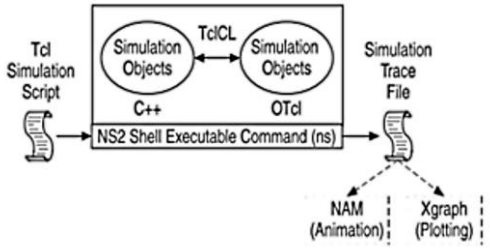


Figure 5 Basic architecture of NS2 Simulator [13].

B. Changes in NS2 code

The main changes were made in the following NS2 C++ functions (in file NS2.35/mac/mac-timer.cc):

1) BackoffTimer::start (int cw,int Idle,double difs)

```
{……
Double slot = mac_phymib_.getSlotTime();
rtime=(Random::random()%cw)*slot;
difs_wait=difs;
If (rtime>= difs) difs_wait=0;        // Added
……
s.schedule(this, &intr, rtime+difs_wait);
}
```

2) BackoffTimer::resume (double difs)

```
{……
Difs_wait = difs;
If  (rtime>= difs) difs_wait=0;        // Added
        ……
s.schedule(this, &intr, rtime+difs_wait); }
```

C. Simulation parameters

Table III.gives the main simulation parameters which remain the same for all simulated scenarios.

CWmin has been adapted to the size of the network; both values 31 and 63 have been used.

TABLE III SIMULATION PARAMETERS

Parameter	Value
MAC protocol	802.11 b (DSSS PHY)
CWmin CWmax	31, 63 1023
Basic rate, DataRate, CBR Rate	1Mb/s, 2Mb/s, 1600Kb/s
Packet Size	512 Byte
Traffic (Flows)	UDP/CBR
Simulation Time, recording period	80s, 2s
Mobility	none
RTSThreshold	3000 bytes (RTS/CTS mechanism disabled)
CSThreshold, RXThreshold (carrier sense and communication ranges),	550m, 250m

6. SIMULATION RESULTS AND EVALUATION

In order to highlight the benefits of our approach, we have simulated it on different scenarios and network topologies.

A. Single-hop topologies

1) Scenarios

Figure 6 presents three single-hop scenarios. The first one (Fig. 5a) contains 3 nodes and 2 flows (a flow from node 0 to node 2 and another one from node 1 to node 2). In order to increase the contention level of the network, new nodes and new flows are added to obtain scenario 2 and 3 (Fig. 5b, Fig.5c). All nodes are within the communication range of each other; they use the same channel and start at the same time.

We consider flow 1 as the main flow and the others as secondary flows. The latter play the role of disruptive flows increasing the interruption probability of the back-offs of flow1 frames.

With these scenarios, we aim at showing that the gain in delay is closely related to the number of back-off interruptions.

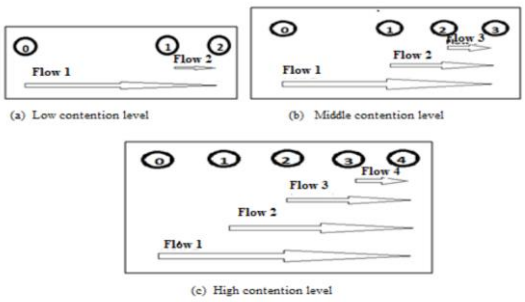


Figure 6 Network scenarios

2) Results, interpretation and evaluation

As expected by the theoretical calculations, the nodes that encounter the most interruptions in their back-offs find their delays the most diminished (more DIFS removed). The nodes that encounter the most interrupts are those that are the furthest away from the receiver (Flow1 for example). The nodes close to the receiver have a higher probability of transmission since their packets arrive in a shorter time at destination. Thus, the delay of flow 1scenario 3 is decreased by 7,24%. The results also show that the greater the number of nodes between the transmitter and the receiver increases, the more the gain of flow1 is important. Finally, the mean end-to-end delay of all scenarios has improved and the average throughput is the same in all cases; however, ORG_DCF remains more efficient when there is no or few contention (flow 3 scenario 2, flows 3 and 4 scenario 3). Table 4 and fig.7 show the simulation results for the single-hop topologies.

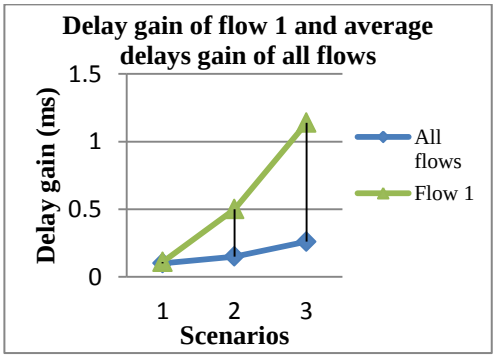


Figure 7 Delay gain flow1 and all flows

TABLE IV SIMULATION RESULTS FOR SINGLE HOP TOPOLOGIES[1]

	Scenario 1				Scenario 2				Scenario 3			
	Delay		*Throughput*		*Delay*		*Throughput*		*Delay*		*Throughput*	
	ORG	DIB	ORG	DIB	ORG	DIB	ORG	DIB	ORG	DIB	ORG	DIB
Flow 1	6,96	6,85	0,64	0,65	10,97	10,47	0,41	0,42	15,73	14,59	0,29	0,30
Flow2	5,83	5,76	0,76	0,76	9,28	9,19	0,48	0,48	15,08	14,82	0,30	0,31
Flow 3					8,80	8,96	0,50	0,50	12,45	12,58	0,36	0,35
Flow 4									10,58	10,80	0,42	0,41
mean	6,40	6,30	0,70	0,70	9,69	9,54	0,46	0,46	13,46	13,20	0,34	0,34
Gain flow1	-0,11		0,01		-0,5		0,01		-1,14		0,01	
Gain rate flow1	-1,58%		1,56%		-4,55%		2,43%		-7,24%		3,44%	

1. Delay (ms), throughput (Mb/s)

B. *Multi-hop topologies*

1) *Scenarios*

Simple two, three and four hops network scenarios have been simulated and compared to a simple (one flow) single hop scenario.

In a multi-hop topology, the destination is out of the sender communication range; the transmitted frame is thus forwarded from node to node till it reaches the destination.

The goal behind these scenarios is to demonstrate that our approach benefits may increase with the number of hops since a same frame will initiate and perform the DCF procedure several times (for each hop).

The four hop network scenario presented in Fig. 8 is based on a string topology with no hidden node problem (all nodes can sense each other). This topology has been inspired from [14].

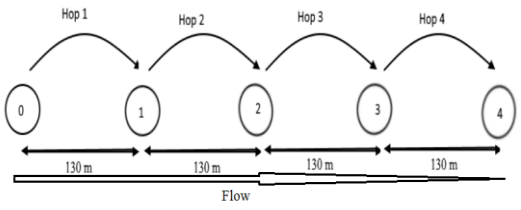


Figure 8 Multi-hop topology

2) *Results, interpretation and evaluation*

Figure 9 shows the results obtained for the four scenarios (one, two, three and four hops). We notice that the delay gain is closely related, this time, to the number of hops.

A very interesting observation is that the difference between the gains relative to two successive scenarios revolves around a DIFS.

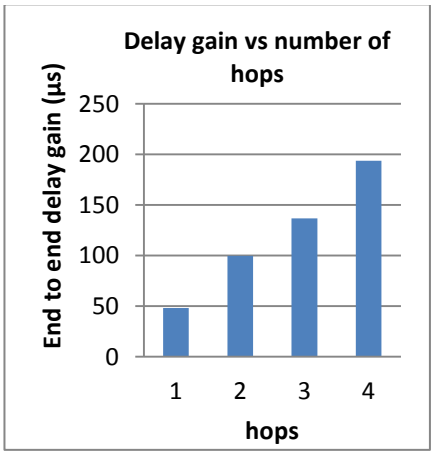


Figure 9 Delay gain versus number of hops

More accurately, it should revolve around p*DIFS; p being the probability for a station to select a back-off number greater than 2 (back-off time greater than DIFS) in the contention window [0,31].

P=29/32 p*DIFS=29/32*50=45,31µs.

This corresponds to the only DIFS removed for each packet transmission (DIFS before back-off start) when back-off time is greater than DIFS.

When a packet is forwarded n times in a multi-hop topology, on average, n*(p*DIFS)are removed from its end to end delay.

There are no DIFS removed before back-off resumes since in these simple scenarios, disruptive flows are absent.

C. *Multi-hop scenarios with disruptive flows*

Single-hop scenarios with different contention levels have allowed us to see the effects of our approach on flows whose backoffs are likely to encounter repeated interruptions (DIFS before backoffs resumes removed when remaining backoff>DIFS).

Simple multi-hop scenarios allowed us to confirm that the delay gain is also proportional to the number of hops. For each hop, the DIFS before backoff start is removed (in case of backoff > DIFS).

Multi-hop scenarios with disruptive flows should combine the benefits of the last two ones.

1) Scenarios

Two hops with three disruptive flows (scenario 1), two with six disruptive flows (scenario 2) and four with three disruptive flows (scenario 3) have been simulated.

Other scenarios have been simulated (four hops with twelve disruptive flows and twenty one nodes for example). However, in these scenarios, flows had very significant access delays (with ORG_DCF and DIB_DCF). It is the well-known IEEE 802.11 ad-hoc unfairness issue where some flows seize completely the channel when others are starved. We thought that the results with such scenarios would be unreliable.

2) Results, interpretation and evaluation

Table V presents the simulation results for the three scenarios mentioned above.

The results are very encouraging, up to 38,38% reduction of delay and 81,08% throughput increase in scenario 2.

TABLE V SIMULATION RESULTS FOR MULTI-HOP TOPOLOGIES[1]

	Scenario 1		Scenario 2		Scenario 3	
	Delay	***Throughput***	***Delay***	***Throughput***	***Delay***	***Throughput***
ORG_DCF	14,81	0,31	115,22	0,037	33,4	0,19
DIB_DCF	13,84	0,32	70,99	0,067	20,98	0,2
Gain	-0,97	0,01	-44,23	0,03	-12,42	0,01
Gain rate	-6,54%	3,22%	-38,38%	81,08%	-37,18%	5,26%

1. Delay (ms), throughput (Mb/s)

7. CONCLUSION AND FUTURE WORK

In this paper, a new approach is proposed to improve 802.11 DCF mechanism; DCF is based on delay principles: Back-off and IFS delays. By noticing that DIFS is, in some cases superfluous, authors have removed it, thus reducing the access time. To study the effects of this new approach, authors tested it on multiple scenarios with increasing contention levels using NS2.35 simulator.The results confirmed theoretical calculations which predicted a diminution of access time equal at least to (n-1) * DIFS (n being the number of interruptions in the back-off process).

By increasing the number of nodes between a transmitter and its destination, we increase the probability that the back-off process of the flow is interrupted by the neighboring flows (flows in the middle); simulation results show that the gain in delay obtained for a specified flow increases proportionally to the number of nodes placed inside the flow.

Thanks to this work, we have identified one of the in depth causes of the well-known IEEE 802.11 ad-hoc unfairness issue. The stations having a large bakoff number are doubly penalized. The waiting time of the backoff is great and these stations are more likely to be interrupted in their backing-off by the stations having smaller backoffs. The more interruptions there are, the more DIFS the stations have to wait.

In a multi-hop topology, for a same frame, DCF procedure is performed several times (each time the frame is relayed by the intermediate nodes). The delay gain is then proportional to the number of hops.

Results obtained with multi hop topologies are very promising. Despite these results, the approach needs to be explored further in order to better understand the new behaviour of the system as a whole and not only for specific flows.

Our contribution focus on IFS and back-off; these two basic mechanisms remain present in many of the amendments that came after IEEE 802.11 b. It would be interesting to simulate it on these ones.

REFERENCES

[1] IEEE Std 802.11, “Part 11:Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” IEEE Computer Society, 2007.

[2] M. Gast, “802.11 wireless networks: The definitive guide,” O'Reilly, 2002.

[3] M. Sapna, “Modeling and Analysis of IEEE 802.11 DCF MAC,” Procedia Computer Science, vol. 57, pp. 473-482, 2015.

[4] D. Dhoutaut, [Free translation]“Study of the IEEE 802.11 standard in ad hoc networks, from simulation to experimentation” “Etude du standard IEEE 802.11 dans le cadre des réseaux ad hoc, de la simulation à l’expérimentation”PhD Thesis, INSA Lyon, pp. 31, 2003.

[5] Z. Hua, L. Ming, C. Imrich, P. Balakrishnan, “A survey of quality of service in IEEE 802.11 networks,”, IEEE Wireless Communications, vol. 11, n° 4, pp. 6-14, 2004.

[6] E. Charfi, L. Chaari, L. Kamoun,“PHY/MAC enhancements and QoS mechanisms for very high throughput WLANs: A survey,”EEE Communications Surveys & Tutorials, vol. 15, n° 4, pp. 1714-1735, 2013.

[7] V. Nitin, D. Anurag, G. Seema, B. Paramvir,“Distributed fair scheduling in a wireless LAN,”IEEE Transactions on Mobile Computing, vol. 4, n° 6, pp. 616-629, 2005.

[8] L. Jinsung, L. Hojin, Y. Yung, C. Song, K. Edward, C. Mung, “Making 802.11 DCF near-optimal: Design, implementation, and evaluation,”IEEE/ACM Transactions on Networking, vol.24, 2016.

[9] B. Vaduvur, D. Alan, S. Scott, Z. Lixia, “MACAW: a media access protocol for wireless LAN's,”ACM SIGCOMM Computer Communication Review, vol. 24, n° 4, pp 212-225, 1994.

[10] B. Swati, K. WU, "An Approach for Improving Performance of Back off Algorithm,"International J. Computer Applications, vol. 46, n° 5, pp. 45-49, 2012.

[11] D. Moad, S. Djahel, F. Nait-Abdesselam, "Padovan sequence based Backoff Algorithm for improved wireless medium access in MANETs,"Global Information Infrastructure and Networking Symposium (GIIS), 2014.

[12] S. Xinghua, D. Lin, "Backoff design for IEEE 802.11 DCF networks: Fundamental tradeoff and design criterion," IEEE/ACM Transactions on Networking (TON), vol. 23, n° 1, pp 300-316, 2015.

[13] T. Issariyakul, E. Hossain, "Introduction to Network Simulator NS2,"Springer Science+Business Media, LLC 2012.

[14] N. Ping Chung, L. Soung Chang, "Throughput analysis of IEEE 802.11 multi-hop adhoc network," IEEE/ACM Transactions on Networking, vol. 15, pp 309-322, 2007.

[15] "The difference between NS2 and NS3," http://www.nsnam.org/support/faq/ns2-ns3, visited 23/02/2018.

Latifa S. Rekik-Mahi obtained her engineering and magister degrees at the University of Oran, Algeria, and She has taught at the computer science department of this same university since 1998. Her first research interests were devoted to the interactive environments of human learning. Currently, she is preparing her "Doctorat" in the area of wireless networks. Her research interests involve scheduling algorithms, routing algorithms and MAC protocols.

Malika Bourenane obtained her "doctorat", magister and engineering degrees at the University of Oran, Algeria, and She has taught at the computer science department of this same university. Her research interests are linked to computer networking, routing, wireless computing, QoS, TCP. She supervises a research team; she contributed through her publications in many scientific international magazines and conference.