ISSN Application No. 21722

Advanced Journal in Wireless and Mobile Communication

Volume No. 11

Issue No. 1

January - April 2023

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

ISSN Application No. 21722

Advanced Journal in Wireless and Mobile Communication

Aims and Scope

Advanced Journal in Wireless and Mobile communication welcomes the original research papers, review papers, experimental investigations, surveys and notes in all areas relating to software engineering and its applications. The following list of sample-topics is by no means to be understood as restricting contributions to the topics mentioned:

•LTE
•4G
•5G
•Ultra wide band communications
•Nnovel mobile applications
•Mobile communications and networking
•Energy-efficient communication networks
•Molecular communications

ISSN Application No. 21722

Advanced Journal in Wireless and Mobile Communication

ISSN Application No. 21722

Advanced Journal in Wireless and Mobile Communication

(Volume No. 11, Issue No. 1, January - April 2023)

Contents

ISSN Application No. 21722

Performance Analysis of MIMO-OFDM for LTE Tactical Communication Systems in Jamming Environment

[1]Settapong Malisuwan, [2]Jesada Sivaraks, [3]Noppadol Tiamnara, [4]Nattakit Suriyakrai

A B S T R A C T

Key challenges for armed forces are the enabling of a highly mobility; full command and control capabilities; critical-infrastructure protection; and leveraging of the commercial, off-the-shelf telecommunications products in military networks. Commercial off-the-shelf (COTS) based solutions serve speed time to market, are easier to maintain, and offer affordable solution when compared to proprietary, government-unique systems. COTS-based LTE are communications technologies, mature, proven reliable and robust, easily deployable, and scalable. However, reliable communications in the COTS-based LTE system are critical in defense applications and the ability to protect jamming signal is key to the tactical communication systems. The objective of this paper is to analyze the performance of the MIMO-OFDM for LTE tactical communication systems in jamming environment. In this research, the average Bit Error Rate (BER) for the 2x1 SFBC 16-QAM and 64-QAM modulation schemes in the asynchronous off-tones (AOTJ) jamming environment with different dynamic ranges of the colored noise jamming is analyzed. MATLAB Monte-Carlo simulations are used to evaluate the performance of this research.

Index Terms— Performance, MIMO-OFDM, LTE, COTS, and Jamming.

I. INTRODUCTION

Armed forces worldwide are transforming from defense-unique systems to COTS-based systems that are affordable to build and maintain. Today, communication systems are readily available and more advanced in the commercial marketplace at reduced costs from a variety of vendors [1]. In this paper, we focus on COTS-based LTE technologies which are mature, proven reliable and robust, easily deployable, and scalable. COTS-based LTE products are reducing costs and increasing connectivity to improve C4I capabilities on the battlefield.

Long Term Evolution (LTE), commonly marketed as 4G LTE, is a wireless high-speed communications networks and portable communications devices such as smartphones, tablets, and laptop computers, and other mobile devices. [2].

The LTE and LTE-A are referred to as System Architecture Evolution (SAE). The main goal of SAE is to provide seamless Internet Protocol (IP) connectivity between the User Equipment (UE) and the Packet Data Network (PDN) with reduced latencies and improved performance using fully optimized for packet-based networks [3]. Fig. 1, provides a high-level view of LTE architecture.

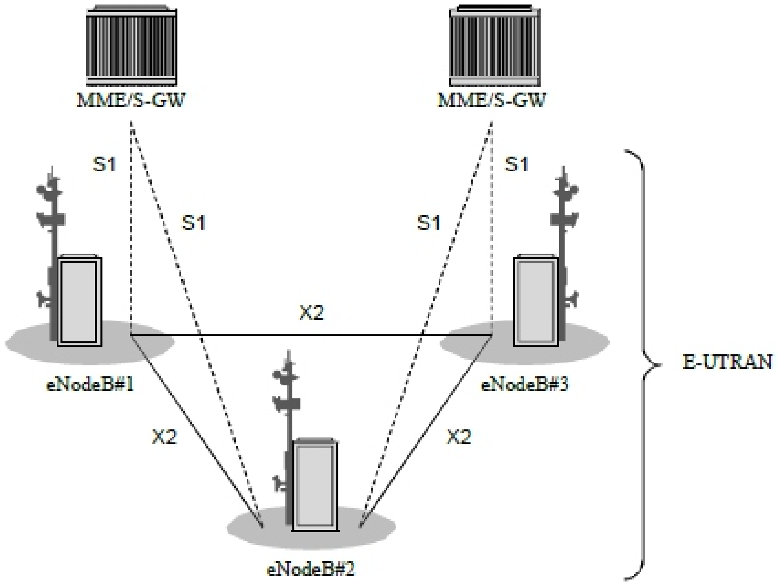


Figure 1. LTE Architecture Overview [4]

The high-level architecture of LTE comprised of three main components, namely the UE, the Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and the Evolved Packet Core (EPC). The E-UTRAN corresponds to the air interface of the network between the UE and the EPC. The Evolved Node B (eNodeB) is the base station for LTE radio without any separate control node and using an X2 interface to communicate with other eNBs, and an S1 interface to communicate with the EPC. This has more flexibility and speed in access during handovers [3].

On the network side, the eNodeB is responsible for the Radio Resource Management (RRM), each of which can be responsible for managing multiple cells. Unlike some of the previous second- and third-generation technologies, LTE integrates radio controller function utilities into eNode B. This minimizes the latency and improving the efficiency between the different protocol layers of the radio access network (RAN). The LTE physical layer is highly efficient in conveying both data and control data between eNode B and mobile user equipment (UE). The simplified block diagram of the LTE downlink physical layer is shown in Fig. 2,.

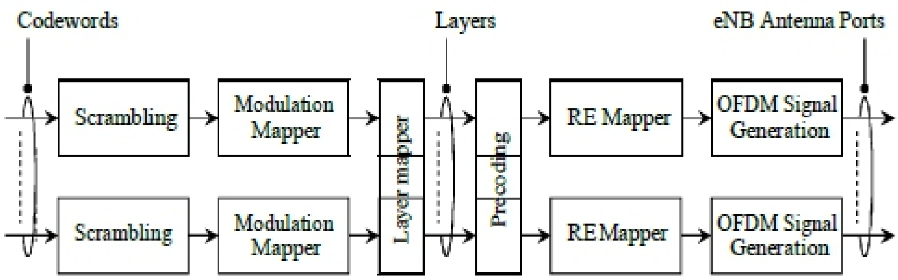


Figure 2. LTE downlink physical layer block diagram [5]

The downlink frame contains the information being sent to the UE that are currently connected to the base station. The scrambling task is performed using a Gold code. The attractive feature of Gold codes is that they can be generated with very low implementation complexity by a simple shift register. Following the scrambling process, the information bits are coded and mapped to complex valued modulation symbols. Once physical channel's codewords have been scrambled and modulated, a layer mapping is applied to the modulated codewords. Next, the layers are precoded using a precoding matrix in 3GPP TS 36.211 which consists of applying coding to the layers of modulated symbols prior to mapping onto Resource Element (RE). After RE mapper operation, the OFDMA mapper combines the precoded values from physical-layer channels. It is the transformation of the complex modulated symbols at the output of RE mapper into a complex valued OFDM signal by means of an Inverse Fast Fourier Transform (IFFT).

In the LTE PHY layer, the Orthogonal frequency division multiplexing (OFDM) and multiple input multiple output (MIMO) are employed to enhance the performance in the LTE network with higher data rates and higher capacity. In addition, the LTE PHY uses Orthogonal Frequency Division Multiple Access (OFDMA) on the downlink (DL) and Single Carrier – Frequency Division Multiple Access (SC-FDMA) on the uplink (UL) [6].

The MIMO-OFDM combination is an attractive air-interface solution which is one of the best way to exploit the spatial diversity, time diversity and frequency diversity. This improves the overall system performance by increasing spectral efficiency to attain throughput of 1 Gbit/sec and beyond, and improving link reliability [7]. Efficient implementation of MIMO-OFDM system is based on the Fast Fourier Transform (FFT) algorithm and MIMO encoding to make efficient use of bandwidth and robustness to multi-path delay. [6]. The MIMO-OFDM systems provide high data rates and are robust to multi-path delay in wireless communications. However, channel parameters are required for diversity combining, coherent detection and decoding [8].

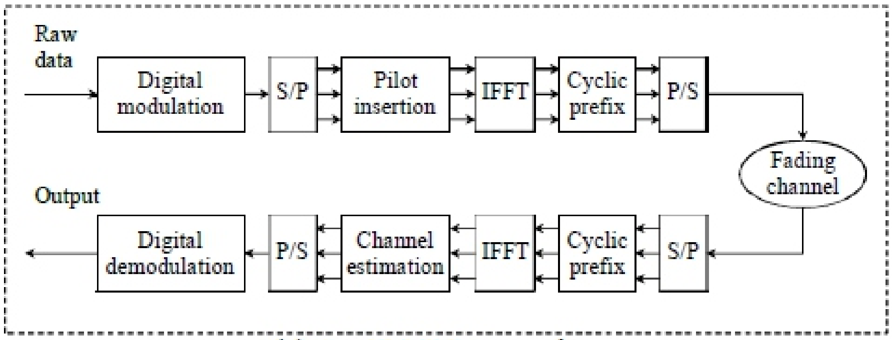
While MIMO-OFDM systems are robust to multipath fading and severe interference, they are not perfect for intentionally jam environments. [9].

In many researches [10], [11], [12], [13], [14], [15], the LTE vulnerabilities by jamming signals is under consideration for military application fields. This drawback is serious concern since it is possible to completely shut down the tactical communication systems running by LTE network with jamming signals. More sophisticated attacks have been discovered as a potentially more effective way to jam LTE networks [16], [17]. Hence, this paper aims to analyze the performance of the MIMO-OFDM for LTE tactical communication systems in jamming environment.

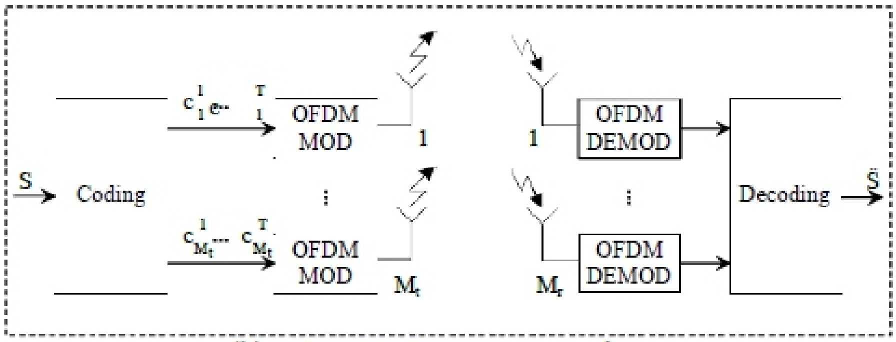
This paper is organized as follows. Section II provides a description of the MIMO-OFDM system model. The Section III presents the proposed model of jamming environment in MIMO-OFDM communication systems. Section IV describes the Bit Error Rate in the LTE MIMO systems. Section V covers the simulations results which are performed in terms of average Bit Error Rate in jamming environment. Finally, the conclusion is specified in the last section.

II. MIMO-OFDM COMMUNICATION SYSTEMS

MIMO systems use the feature of spatial diversity by using spatially separated antennas in a dense multipath fading environment to obtain diversity gain or capacity gain. Advanced techniques in MIMO make a significant increase in performance for OFDM systems with bandwidth efficiencies on the order of 10 b/s/Hz.



(a) An OFDM transceiver



(b) A MIMO-OFDM transceiver

Figure 3. A simplified block diagram of MIMO-OFDM system, where S = [s1, s2, , sNs] denotes a block of Ns data symbols [18].

Fig. 3(a), shows a simplified block diagram of an N-tone OFDM system. First, the incoming data stream is mapped into some modulation scheme such as QPSK or QAM. [19]. After going through all processes, the data symbols are detected with the estimated channel information and the transmitted bit stream is recovered.

A general MIMO-OFDM system is shown in Fig. 3(b), where Mt transmit antennas, Mr receive antennas, and N-tone OFDM are used [19]. First, the incoming bit stream is mapped into a number of data symbols via some modulation type such as QAM. Then a block of Ns data symbols S = [s1, s2, , sNc] are encoded into a codeword matrix C of size NT × Mt, which will then be sent through Mt antennas in T OFDM blocks, each block consisting of N subchannels. Specifically, cj1, cj2, , cjT will be transmitted from the jth transmit antenna in OFDM blocks 1, 2, , T , respectively, where cn denotes a vector of length N, for all j = 1, 2, , Mt and n = 1, 2, , T . The codeword matrix C can be expressed as

$$C = \begin{pmatrix} c_1^1 & \cdots & c_{M_t}^1 \\ \vdots & \ddots & \vdots \\ c_1^T & \cdots & c_{M_t}^T \end{pmatrix} \quad (1)\ [18]$$

After appending the cyclic prefix on each OFDM block, cn will be transmitted from the jth transmit antenna in the nth OFDM block and passed through the MIMO channels. Then, the received signals will be sent to the reverse OFDM block and sent to the decoder.

In LTE, transmit diversity is an effective technique for combating fading by using Space Frequency Block Coding (SFBC). SFBC provides both spatial and frequency diversity and improves cell coverage and/or improves cell-edge throughput. SFBC is a frequency domain adaptation of renowned Space-time Block Coding (STBC). STBC is also recognized as Alamouti coding [20].

The advantage of SFBC over STBC is that in SFBC coding is done across the sub- carriers within the interval of OFD OFDM symbol while STBC applies coding across the number of OFDM symbols equivalent to number of transmit antennas [20]. Fig. 4, illustrates the SFBC operation for the particular two-antenna configuration. The transmitters send the same underlying user data, but in different parts of the RF frequency space.

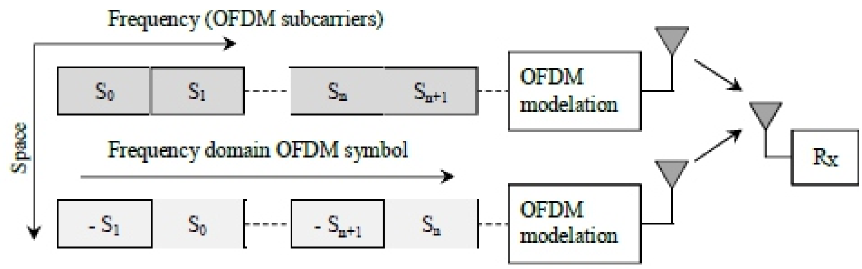


Figure 4. Space Frequency Block Coding SFBC assuming two antennas [21]

III. LTE MIMO-OFDM SYSTEM IN JAMMING ENVIRONMENT

As any kind of wireless network systems, LTE is also vulnerable to radio jamming attacks especially in the case of next-generation COTS tactical communication systems. A simple method for radio jamming is the transmission of radio signals to disrupt communications by decreasing the Signal-to-Noise ratio (SNR) of the received signal. This jamming transmits a high-power signal over the entire target bandwidth of the victim system [22].

Jamming is different from network interferences because it describes the tactical use of electromagnetic signals in an intent to disrupt communications but interference is unintentional forms of disruptions. Intentional interference or jamming is usually operated by an attacker who intends to interrupt communications within or between wireless networks. Different techniques of jamming attacks can be conducted, from hindering transmission to distorting packets in legitimate wireless communications.

Tactical LTE communication systems should be able to operate in spectral environments fraught with interference and jamming. The wireless channels are subject to attack from jamming signals. This causes the performance of the network to degrade. In this paper, the Asynchronous Off-Tone Jamming (AOTJ) is focused.

There are two types of Asynchronous Off-Tone Jamming (AOTJ). The first type is called single off-tone jamming and the second type is multiple off-tone jamming attack. The operational concept of this technique is to transmit asynchronous off-tones which generates inter channel interference (ICI) of the OFDM signal at the receiver [23]. Also the side-lobes of the signal not aligned with the orthogonal OFDM subcarriers due to frequency offset will create non-zero components at the sampling period that can be a source of ICI. AOTJ is efficient and practical for attackers because the jamming signal does not need frequency matching with target signal. The example of the two types of AOTJ can be seen in Fig. 5,.

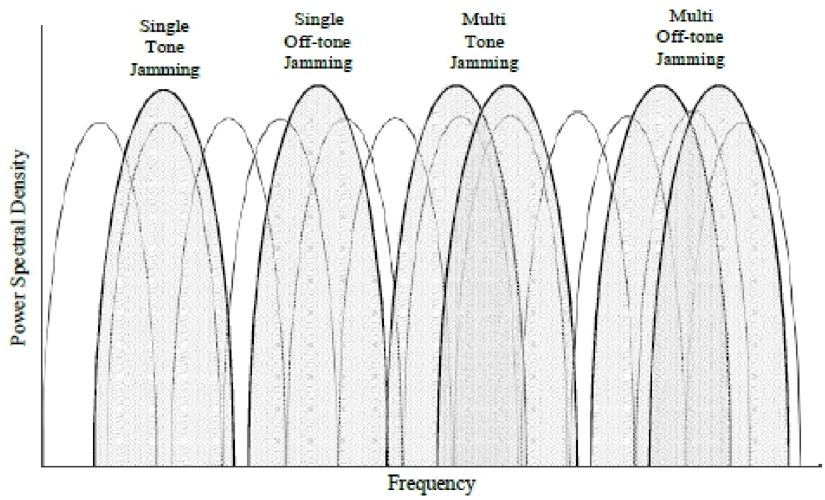


Figure 5. Different jamming attacks on LTE downlink [23]

IV. BER ANALYSIS OF LTE MIMO SYSTEM

In a MIMO system with Nr receive antennas and Nt transmit antennas, the relation between the received and the transmitted signals on OFDM subcarrier frequency k (k ∈ 1, ..., N), at sampling instant time n is given by

$$y_{k,n} = H_{k,n}x_{k,n} + n_{k,n} \tag{2}$$

where $y_{k,n} \in C_{Nr\times 1}$ is the received output vector, $H_{k,n} \in C_{Nr\times Nt}$ represents the channel matrix on subcarrier k at instant time n , $x_{k,n} \in C_{Nt\times 1}$ is the transmit symbol vector and $n_{k,n} \sim CN(0, o_n^2. I)$ is a white, complex valued Gaussian noise vector with variance o_n^2 and I is an $N_r \times N_r$ identity matrix.

Assuming perfect channel estimation, the channel matrix and noise variance are considered to be known at the receiver. A linear equalizer filter given by a matrix $F_{k,n} \in C_{Nr\times Nr}$ is applied on the received symbol vector $y_{k,n}$ to determine the post-equalization symbol vector $r_{k,n}$ as follows [24]

$$r_{k,n} = F_{k,n}y_{k,n} = F_{k,n}H_{k,n}x_{k,n} + F_{k,n}n_{k,n}. \tag{3}$$

The Zero Forcing (ZF) or Minimum Mean Square Error (MMSE) design criterion is typically used for the linear receiver and the input signal vector is normalized to unit power [25]. In MIMO-OFDM systems, the key factor of link error prediction and performances is the signal to noise ratio (SNR) which represents the measurement for the channel quality information. In this study, the SNR is defined as follows [26]:

$$\gamma_{k,n} = \frac{\left\| H_{k,n}x_{k,n} \right\|_F^2}{N_t\,\sigma_n^2} \tag{4}$$

where xk,n is the transmitted symbol vector, ‖. ‖2 is the squared Frobenius norm of a matrix.

Average BER Performance analysis for several M-QAM Schemes

In this section, a Bit Error Rate (BER) analysis is presented for Multiple-Input Multiple Output (MIMO) schemes in the 3GPP Long Term Evolution (LTE) system. The average BER of the system is analyzed over flat Rayleigh fading channels by applying M-ary quadrature amplitude modulation (M-QAM) schemes. The analysis is based on the probability density function of the instantaneous Signal to Noise Ratio and the Moment generating function.

Analysis for 2 × 1 SFBC-OFDM Scheme

When two eNodeB antennas are available for transmit diversity operation, the Space Frequency Block Code (SFBC) is used [27]. SFBC is based on the well known Space Time Block Code (STBC), derived by Alamouti for two transmit antennas [28]. In LTE, for SFBC transmission, the symbols are transmitted from two eNodeB antenna ports on each pair of adjacent subcarriers as follows [27]:

$$\begin{bmatrix} y^{(0)}(1) & y^{(0)}(2) \\ y^{(1)}(1) & y^{(1)}(2) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \tag{5}$$

where y(p)(k) denotes the symbols transmitted on the kth subcarrier from antenna port p . An important characteristic of such codes is that the transmitted signal streams are orthogonal and a simple linear receiver is required for optimal performance.

This paper applied the BER expressions over flat Rayleigh fading channels, given by Pb (E) from the reference [2]. Then, the overall average BER over N subcarriers, in each case can be calculated from

$$BER = \frac{1}{N}\sum_{k=1}^{N} P_{b,k}(E) \tag{6}$$

where the index k is the subcarrier index.

For the 2 × 1 SFBC MIMO scheme, the probability density function of the SNR for each subcarrier is f(y).

$$f(\gamma) = \frac{2}{\bar{\gamma}^2}\gamma e^{-\frac{2}{\bar{\gamma}}\gamma} \tag{7}$$

To derive the BER, we follow the unified approach to the performance analysis of digital communication systems over generalized fading channel [29]. To this end, we first derive the expression of Moment Generating Function (MGF) of the derived probability density function of the instantaneous SNR as [30]:

$$M_{\bar{\gamma}}(s) = \int_0^{\infty} e^{-s\gamma} f(\gamma)d\gamma. \tag{8}$$

The average BER expression for M-QAM modulation scheme can be obtained from [30] as:

$$P_b(E) \cong B \sum_{i=1}^{\sqrt{M}/2} \frac{1}{\pi}\int_0^{\pi/2} M_{\bar{\gamma}}\left(A_{i,\theta}\right)d\theta \tag{9}$$

As described in Section III, the AOTJ Jamming type is considered in this research. In this scenario, this paper used the jamming model from [31].

To describe the concept of the jamming signal model we restate the derivation of the model from the reference [32] in this section.

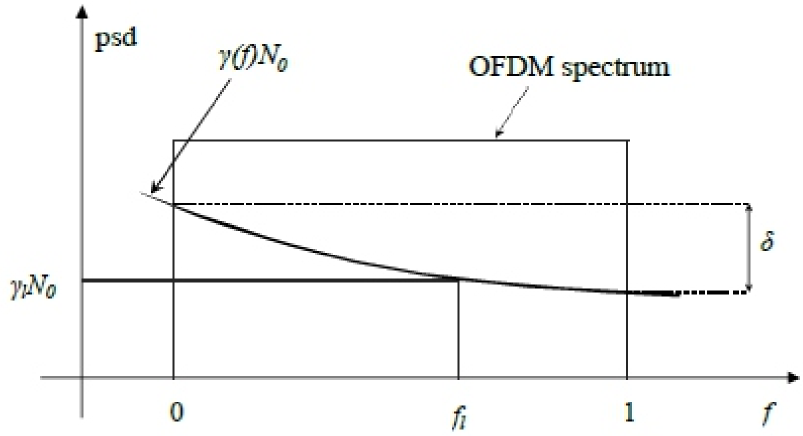


Figure 6. An OFDM user spectrum and the psd of the colored noise process over the normalized frequency.

In our jamming signal model, we adopt the concept of the colored noise process over the normalized frequency in [32]. Fig. 6, shows the illustration of an OFDM user spectrum and the psd of the jamming signal over the normalized frequency. ð is the dynamic range, defined as:

$$\delta = \frac{\gamma_{max}}{\gamma_{min}}. \qquad (10)$$

It is the power ratio between the highest and lowest value of y.

To model the shape of its psd, we define a frequency dependent power weighting factor y ∈ ℝ+ . So the noise psd at frequency ƒS is ySN0 with yS = y(ƒS). For convenience, we normalize the frequency so that ƒ = 0 represents the left edge and ƒ = 1 represents the right edge of the OFDM bandwidth. A symbol that is transmitted at frequency ƒS , is then distorted by AWGN with noise power spectral density ySN0.

Based on the reference [33], the frequency interleaver maps every complex symbol xi to a certain frequency ƒi. So if we assume ideal interleaveing, this frequency can be regarded as random variable, that is uniformly distributed over the interval (0, 1).

This leads to our proposed discrete channel model in Fig. 7,. The sequence x of K complex data symbols is distorted by additive noise n = (n1, … , nK)T . This noise vector n results from the multiplication of white Gaussian noise w = (w1 , … , wK)T with variance E{w2} = o2 = N0⁄2 and the matrix of weighting factors L = diag(√ y1, … , √ yK). The factors ƒyi can be found via the transformation of the uniformly distributed random variable ƒ through yi = y(ƒi). The input to the receiver is

$$y = x + Lw \qquad (11)$$

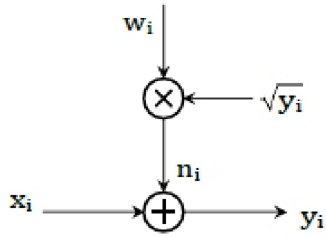


Figure 7. Proposed channel model for additive colored Gaussian noise

V. SIMULATION RESULT AND DISCUSSIONS

The aim of this research is to analyze the average Bit Error Rate (BER) for the 2x1 SFBC 16-QAM and 64-QAM modulation schemes in the asynchronous off-tones (AOTJ) jamming environment with different dynamic ranges of the colored noise jamming. In our study, jamming signal is transmitted asynchronous off-tones which are not perfectly periodic or have an offset at the sampling frequencies. Thus, it creates inter channel interference (ICI) of the OFDM signal at the receiver.

Table I: Simulation Parameter Setting

Parameter	Setting
Transmission Schemes	2×1 SFBC
Bandwidth	5MHz
Simulation length	5000 subframes
Channel Type	Flat Rayleigh
Channel knowledge	Perfect
CQI	9(16-QAM) and 16(64-QAM)
Dynamic range (δ)	δ −2dB, 10dB

The average BER performance as a function of E_c/N_0 for 2×1 SFBC with different modulation modes has been analyzed.

Fig. 8, and Fig. 9, show the simulation results for 16-QAM and 64-QAM modulations respectively

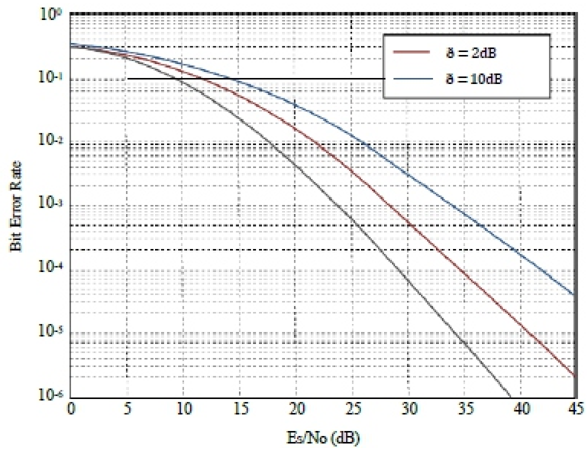


Figure 8. Monte-Carlo Simulation of the average BER for 16-QAM modulation

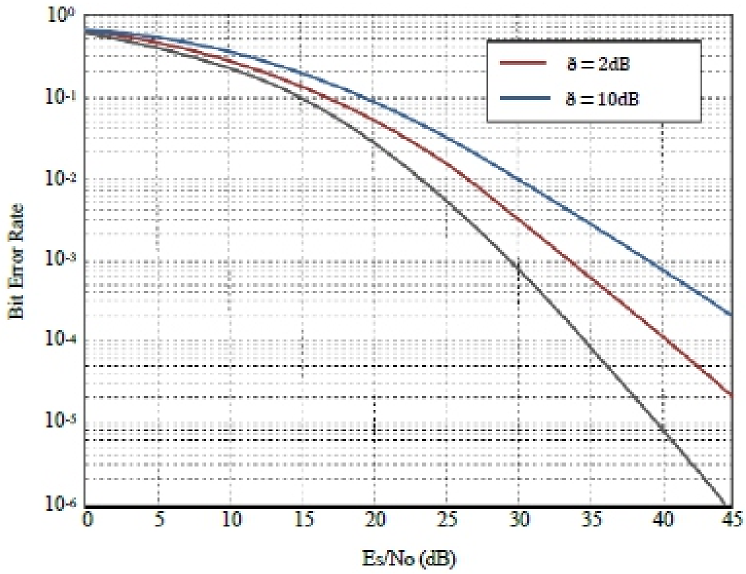


Figure 9. Monte-Carlo Simulation of the average BER for 64-QAM modulation

Fig. 8, and Fig. 9, show the average BER for different dynamic ranges versus Ec⁄N0 that is simulated by MATLAB. This result is indicated that when increases the dynamic ranges of the colored noise jamming, the effect on the reducing Ec⁄N0 exhibits increasing average BER. Equivalently, the increasing in the noise jamming in this study will create higher interchannel interference (ICI) of the OFDM signal at the receiver.

CONCLUSION

In this paper we analyzed the performance of LTE for the 2x1 SFBC 16-QAM and 64-QAM modulation schemes in the asynchronous off-tones (AOTJ) jamming environment. Monte-Carlo simulation is used to demonstrate the performance of the LTE in term of Bit Error Rate. The simulation results in this research are indicated that LTE is extremely vulnerable to adversarial jamming. This is not surprising results, considering LTE was not designed to be a military communication system. Therefore, the commercial LTE technology needs to be developed to utilize in tactical environment to match with the military requirement.

REFERENCES

[1] Harry Jensen and Jeff Sharpe, "LTE Infrastructure for Command and Control Networks", Radisys, June 2013 : https://embedded.communities.intel.com/servlet/JiveServlet/ previewBody/7764-102-3-3098/LTE_C2 _Radisys_WP.pdf

[2] Understanding 3GPP Release 12 Standards for HSPA+ and LTE-Advanced Enhancements, 4G Americas publication, Feb 2015.

[3] Long Term Evolution Protocol Overview, White paper, Freescale Semiconductor, Inc. 2008.

[4] http://www.cse.unt.edu/~rdantu/FALL_2013_WIRELESS_N ETWORKS/LTE_Alcatel_White_Paper.pdf

[5] http://publications.polymtl.ca/1333/1/2013_AliJemmali.pdf

[6] A. Omri and R. Bouallegue, "New Transmission Scheme for MIMO-OFDM System", International Journal of Next-Generation Networks (IJNGN) Vol.3, No.1, March 2011.

[7] J. Ketonen, M. Juntti and J. R. Cavallaro, " Performance Complexity Comparison of Receivers for a LTE MIMO–OFDM System ", IEEE Transaction on Signal Processing, Vol. 58, No. 6, pp. 3360 – 3372, June 2010.
[8] Hussein Hijazi, Eric Pierre Simon, Martine Lienard, Laurent Ros, "Channel Estimation for MIMO- OFDM Systems in Fast Time-Varying Environments", Proceedings of the 4th International Symposium on Communications,Control and Signal Processing, ISCCSP 2010, Limassol, Cyprus, 3-5 March 2010.
[9] M. Wilhelm, I. Martinovic, J. B. Schmitt, and V. Lenders, "Reactive jamming in wireless networks - how realistic is the threat?" in Proc. of WiSec, June 2011.
[10] M. Lichtman, J. Reed, M. Norton, T. Clancy, "Vulnerability of LTE to Hostile Interference", IEEE Global Conference on Signal and Information Processing (GlobalSIP), Dec 2013.
[11] T. Clancy, "Efficient OFDM denial: Pilot jamming and pilot nulling," in Communications (ICC), 2011 IEEE International Conference on, June 2011.
[12] J. Luo, J. Andrian, and C. Zhou, "Bit error rate analysis of jamming for OFDM systems," in Wireless Telecommunications Symposium, 2007. WTS 2007, April 2007.
[13] S. Chao, W. Ping, and S. Guozhong, "Performance of OFDM in the presence of multitone jamming," in Robotics and Applications (ISRA), 2012 IEEE Symposium on, pp. 118 –121, June 2012.
[14] M. Han, T. Yu, J. Kim, K. Kwak, S. Lee, S. Han, and D. Hong, "OFDM channel estimation with jammed pilot detector under narrow- band jamming," IEEE Transactions on Vehicular Technology, vol. 57, no. 3, pp. 1934–1939, 2008.
[15] T. Clancy, "Efficient OFDM denial: Pilot jamming and pilot nulling," in Proc. of ICC, 2011.
[16] D Talbot, One simple trick could disable a city 4G phone network. (MIT Technology Review, 2012). http://goo.gl/jROMe2.
[17] M. Lichtman, J.H. Reed, T.C. Clancy, M. Norton, "Vulnerability of LTE to HostileInterference", in Proceedings of the IEEE Global Conference on Signal and Information Processing, GlobalSIP '13, Austin, TX (IEEE New York, 2013), pp. 285–288.
[18] http://www.eecis.udel.edu/~xxia/WeiZhang1.pdf
[19] Wei Zhang, Xiang-Gen Xia and Khaled Ben Letajef, "Space-Time Frequency Coding for MIMO-OFDM in Next Generation Broadband Wireless Systems", IEEE Wireless Communications, June 2007.
[20] M. I. Rahman, N. Marchetti, S. S. Das, Frank H.P. Fitzek and Ramjee Prasad, "Combining Orthogonal Space-Frequency Block Coding and Spatial Multiplexing in MIMO-OFDM System," Center for TeleInFrastruktur (CTiF), Aalborg University, Denmark.
[21] http://www.bth.se/fou/cuppsats.nsf/all/20c60f89a2ad50f3c12 5764b0048aaf4/$file/thesis_report_MEE0 958.pdf
[22] W. Xu, Y. Zhang and T. Wood, "The feasibility of Launching and Detecting Jamming Attacks in Wireless Networks", in ACM MOBIHOC; Urbana-Champaign (ACM New York, 2005), 2005.
[23] Chowdhury Shariar, Shabnam Sodagari, Robert McGwier and T. Charles Clancy, "Performance Impact of Asynchronous Off-tone Jamming Attacks against OFDM" ICC, page 2177-2182. IEEE, 2013.
[24] SCHWARZ, S., MEHLFUHRER, C. and RUPP, M. (2010). Calculation of the spatial preprocessing and link adaption feedback for 3gpp umts/lte. Wireless Advanced (WiAD), 2010 6th Conference on. 1 –6.
[25] TSE, D. and VISWANATH, P. (2008). Fundamnetals of Wireless Communications, Cambridge University Press.
[26] MEHLF¨UHRER, C., WRULICH, M., IKUNO, J. C., BOSANSKA, D. and RUPP, M. (2009). Simulating the long term evolution physical layer. Proc. of the 17th European Signal Processing Conference (EUSIPCO 2009). Glasgow, Scotland.
[27] SESIA, S., ISSAM, T. and BACKER, M. (2011). LTE The UMTS Long Term Evolution From Theory To Practice, John Wiley.
[28] ALAMOUTI, S. (1998). A simple transmit diversity technique for wireless communications. Selected Areas in Communications, IEEE Journal on, 16, 1451 –1458.
[29] SIMON, M. K. and ALOUINI, M.-S. (1998). A unified approach to the performance analysis of digital communication over generalized fading channels. Proceedings of the IEEE, 86, 1860–1877.
[30] ALOUINI, M. S. and SIMON, M. K. (2000). Digital Communications over Fading Channels: A Unified Approach to Performance Analysis, Wiley.
[31] S. Malisuwan, J. Sivaraks, N. Madan and N. Suriyakrai, "Performance Analysis of LTE Multicast Systems in the Presence of the Colored Noise Jamming," International Journal of Information and Electronics Engineering, Vol. 5, No. 1, January 2015.
[32] F. Klingler, H. Schulze, "Adjacent Channel Interference Effects in OFDM Systems with Imperfect Anti-Aliasing Filtering," South Westphalia University of Applied Sciences, Campus Meschede.
[33] F. Klingler, H. Schulze, "Adjacent Channel Interference Effects in OFDM Systems with Imperfect Anti-Aliasing Filtering" South Westphalia University of Applied Sciences, Campus Meschede.

ISSN Application No. 21722

Resolution of Geographical Names in Information Extraction

Amal Alshahrani
University of Manchester, School of Computer Science
amal.alshahrani@gmail.com

ABSTRACT

This study is concerned with using a machine learning method for extracting geographical named entities from the unstructured text and using a GeoNames database to differentiate between place and non-place references from the recognized geospatial named entities. It also resolves ambiguous place names (e.g. a place name referring to many places) into unambiguous place references using contextual clues and other assumptions. The GENIA tagger is used to annotate the text, and the YamCha is used to train and test the data sets. To evaluate our results, we have also used the CAFETIERE system on the test data sets. The evaluation process shows that our results have a high recall and has an approximately equivalent F-measure with the evaluation results confirming the validity of the chosen method.

I. INTRODUCTION

Nowadays, the internet is a major source of knowledge for humans. The amount of data available on the Internet is exponentially increasing day by day. The user can easily get data from any website on the Internet. For example, he can obtain data from Wikipedia as web pages, and he can get another data from scientific websites (e.g. Geojournal) as electronic papers. There is a need for a technology or a tool to analyse the data obtained from the Internet, get information, and acquire knowledge from this data. This information and knowledge can be later used for different purposes.

Information extraction[1][2] is the process of searching a text for information. There are two main approaches for designing an information extraction system, knowledge engineering, and automatic training. Each of these two approaches has its own advantages and disadvantages and is used based on the resources available to a designer of the system. In this research, we are going to follow the automatic training approach to address the problem of extraction and resolution of geographical names from the unstructured English text. The geospatial data is very important nowadays for many purposes.

For example, Geo Names website is gaining popularity for helping people to search and locate geographical entities over the web. However, some places occur as common words, or they exist repeatedly across the globe. If someone typed in, for instance, the word "Research" in http://www.geoNames.org, he/she'll find Research Australia, Victoria populated place S 37° 42' 0" E 145° 11' 0"). He/she also will find that many place names occur repeatedly across the globe. For many applications, it is important to know (1) whether a string is a place name reference at all, and (2) if it is the case, then which specific place it represents?

1. Background

Two approaches are used to extract information from the text: knowledge engineering approach, an automatic training approach. These approaches are described in the following sections.

Knowledge Engineering Approach[3]: In the knowledge engineering approach, the information extraction is performed by manually constructing and writing a set of rules. This approach is also called the rule-based approach. Typically, information extraction rules are developed by a domain expert, called a "knowledge engineer". On the other hand, the designer should be familiar with the formalism of the target system to be used. Usually, the knowledge engineer identifies texts of interest from the selected domain and then the designer identifies common patterns by using her or his intuition to develop the corresponding rules. The rules are then implemented in the information extraction system which interprets and uses them to extract useful facts from the given text. The rule-basedapproach for extracting information requires time and effort to manually develop the rules and check them against a number of texts in order to verify the correct development of rules for the desired results.

Automatic Training Approach[3]: The automatic training approach for information extraction, also called the "machine learning approach", does not involve the manual design and development of the rules. Instead, this approach uses statistical methods which automatically extract rules from the texts given as the training data. However, a large amount of training data is needed to accurately identify the rules. In addition, manual annotation of the training data is required before executing the algorithm. This typeof learning is also called "supervised machine learning".The automatic training approach includes algorithms such as decision trees, Hidden Markov Models (HMM) [4], Condition Random Fields (CRF) [5][6][7], and Support Vector Machines (SVM) [8].

II.DETAILS EXPERIMENTAL

3.1 Methodology

The methodology comprises four stages:

- **Collecting data and preprocessing:** In this stage, unnecessary data from each article will be manually removed. We have built a corpus consisting of articles taken from Geojournal[1] and Wikipedia[2].

[1] *http://www.springer.com/social+sciences/population+stu dies/journal/10708*
[2] *http://en.wikipedia.org/wiki/Wikipedia*

- **Text annotations (Annotator)**: The objective of the Pre-processing stage is to clean the articles which will make our corpus. This pre-processing is done manually for each article. The objective of the Annotator is to split the input article into word, POS, Chunk and NEs tags. To achieve this object, the GENIA tagger[3] has been used.

- **GeoExtractor component:** The objective of the GeoExtractor is to use the output of the Annotator to generate the candidate GeoNames. YamCha tool[4] has been used to achieve this objective. Then, a GeoNames database (e.g. CAFETIER database) is fed from the output of the YamCha to confirm the GeoNames obtained by YamCha[9].

- **GeoResolver component:** The goal of the GeoResolver is to resolve the geographical ambiguity of the GeoName extracted from the GeoExtractor (i.e. To differentiate between ambiguous and non-ambiguous GeoNames). One way to accomplish this is to use the context surrounding each GeoNames and other assumptions.

This study focused on extracting geospatial NEs from the unstructured text. To train our data, we have built a corpus consisting of articles taken from Geojournal and Wikipedia. We have chosen Wikipedia and Geojournal because they are easy to access and to use. As they contain many types of articles, we only focused on the ones containing lots of geospatial location names (e.g. states, and cities' names). From Wikipedia, we selected 30 articles written about wars and battles because they are full of geospatial names. From Geojournal, we have selected 30 articles about GeoHealth (EthnoBiology and EthnoMedicine) which contain many of GeoNames. Prior to the Annotation process, since we only focus on the text of each article, we need to remove the specific mark-up signs, and other information that does not relate to the aim of our project.

3.2 Pre-process of the corpus Preprocessing steps of Geojournal articles:

To prepare the Geojoural article the following steps should have followed: Removing header-note and footnote, Removing an article title, Removing authors' names and their affiliations, Removing figures, images and tables, and Removing references.

---

[3] *http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/tagger/*
[4] *http://chasen.org/~taku/software/yamcha/*

Preprocessing steps of Wikipedia articles:

To prepare the Wikipedia article the following steps should have followed: Removing the end of each article, Removing HTML markup from the page, Removing info boxes in the Wikipedia pages, Removing Wikipedia-format links, and Removing special symbols.

3.3 Guidelines of Annotation:

To annotate our data set, we have followed a number of guidelines[5]. These guidelines are described below.

1. Continent: Any continent name (Asia, South America, Australia, Europe, North America, Africa, and Antarctica)

2. Nation: Any nation entity (e.g. UK, USA, KSA, U.S and Soviet Union... etc.)

3. State-or-Province: A word representing a state and a province is tagged by LOC.

4. County-or-District: Any word representing a county or a district is tagged.

5. Cluster: Will annotate the entities representing a group of countries. Examples of these entities include Eastern Europe, the European Union, the Middle East, Southeast Asia, and Latin America.

6. Nested Region Names: Nested region names will be tagged as one entity (i.e. location) per each region.

7. Ref. Location: Annotate the entities that refer to a geographic position. We annotate to the location from this type as illustrated by the following example: where the parent is always a country, and the child is always a city, village, or a province in this country.

8. Land-Region-natural: We annotate the entities that related to the non-artificial locations such as Caucasus.

9. Region-International: The entities that cross the borders of the national have been tagged.

---

[5] *http://projects.ldc.upenn.edu/ace/annotation/*

10. River, valley, and lake: Any river, valley, or lake name is tagged.

There are some vague, which should be avoided while selecting the Annotation method. This is explained below[6].

Military words: such as, Iraqi troops

People words: such as, Chinesepeople

Government words: such as,Russiangovernment

Organization words: such as, U.S.Fish and Wildlife Service.

3.4 Annotation process

We have used GENIA tagger to annotate the data, The GENIA tagger is not designed for annotatingGeoNames. GENIA tagger is a tool which takes English text as an input and produces a table consist of four columns: word, POS (Part-Of- Speech)[7], Chunk and NEs tag. GENIA annotates the data by splitting the sentence into words.Each word then is tagged either "O: out of named entities", or other symbols related to biomedical NEs. To use it with the GeoNames, we manually tagged each place name to "LOC". The symbol "O" indicates that the word is not part of a geospatial NEs, and "LOC" indicates that the word is part of a geographical NEs.

III. RESULTS AND DISCUSSION

In addition to training and testing our system on Wiki and Geojo, we have conducted several tests on the training data set. We have used the following corpuses, Geojournal and Wikipedia, Reuters[8] (RC), CoNLL2003[9], GeoNames database to further train our corpus. As shown in Table 1, a combination of corpora has been used for this train. These combinations were tested under a YamCha tool. From the results shown in Table 1 the following remarks can be drawn.

[6] *http://projects.ldc.upenn.edu/ace/docs/English-Entities- Guidelines_v6.6.pdf*
[7] *Part-of-Speech tagging: the process of marking up the words in a text as corresponding to a particular part-of-speech, based on both its definition, as well as its context. See Appendix A.*
[8] *http://trec.nist.gov/data/reuters/reuters.html.*
[9] *http://www.cnts.ua.ac.be/conll2003/ner/*

Table 1: Results of Corpus combination

Corpus	Precision	Recall	F-measure
Geojo	0.50	0.63	0.56
Wiki	0.66	0.47	0.55
Geojo + Wiki	0.56	0.53	0.54
RC + CoNLL03 + GeoDB	0.50	0.89	0.64
GeoDB+RC+CoNLL03+GeoJo+wiki	0.60	0.80	0.69

Firstly, when using Geojo, we have found that the recall is higher than using Wiki'scorpus. On the other hand, the precision with Wiki is higher than the precision with Geojo. Secondly, when we have added Wiki to Geojo, we have got a balance in both the precision and the recall comparing to the individual results. Thirdly, when we used a corpus consisting of RC, CoNLL03, and GeoDB, we noticed that the recall is the highest result comparing with the other combination results, we have obtained 89%.The last but not the least, when we added (Geojo and Wiki) to the corpus of RC.

CoNLL03, and GeoDB, we found that there is a trade-off between the recall and the precision comparing with the corpus consisting of RC, CoNLL03, and GeoDB. Thus, we have got the best performance F-measure 0.69%. Comparing with Witmer's work [10], the result of our F-measure is better than Witmer‟s result. The latter has got 0.67%, while we have got 0.69%.

2. EVALUATION

We are using CAFTETIER system10 which is used to evaluate our results and perform looking-up in GeoNames database to separate actual place references and non-place references. From the comparison made between the results of YamCha and CAFETIER, as shown in Table 2, the following remarks can be drawn. Firstly, the precision is improved by %11 using CAFTETIER over YamCha. The main reason for this improvement is that CAFTETIER system contains a GeoNames database which helps the CAFTETIER system to extract the accurate entities. On the other hand, YamCha does not come with any geospatial database for the place names.Secondly, the recall result of CAFETIER system is less by %10 than the results of YamCha. This is because YamCha is given manually annotated data for any place names, regardlessits semantic meaning (i.e. Whether this place name refers to an organisation or a person's name, etc.). This allowed our system to retrieve a high volume of location names.Thirdly, the F-measure of both CAFETIRER and YamCha systems is approximately the same. This is due to the remark that the precision is increased and the recall is decreased in the CAFETIER results, whereas the precision was low and the recall was high in the YamCha results.

[10] *http://www.nactem.ac.uk/cafetiere/*

Table 2:Comarision between CAFETIER and YamCha

	Precision	Recall	F-Measure
YamCha Result	0.60	0.80	0.69
CAFETIER Result	0.71	0.70	0.67

CONCLUSIONS

The main aim of this project was to design and develop a software component based on the supervised machine learning algorithm to extract geographical named entities from the unstructured text. To achieve this aim, we have accomplished the following objectives. Firstly, we have investigated the approaches of supervised classification that could be used to deal with geographical entities in terms of both performance and accuracy. Thus, we have chosen the automatic training approach. This is because this approach does not need an expert knowledge engineer to annotate the text. Secondly, we have designed our system. This system consists of three stages. The first stage (i.e. Annotator); we have used a tool called GENIA tagger to annotate this corpus. Then, we manually annotated the GeoNames in the output of the GENIA tagger and chosen the word-features by which the GeoNames will be selected. In the second stage, called GeoExtractor, we have used a machine learning classifier, called YamCha, to train and to test our corpus.Then, YamCha is tested on a corpus of 60 articles taken from Wikipedia and Geojouranl. The overall results obtained are as follows: precision of 60% and a recall of 80% and F-Measure of 69%. Finally, we have evaluated the test results obtained from YamCha. To do so, we first examined a system called CAFETIERE and then used it to evaluate our results. This evaluation has shown approximately equivalent results to our obtained results.

ACKNOWLEDGMENTS

Author would like to express their deepest appreciation to Manchester university andculture bureau of saudiarabia in uk, which is supported by the government of the kingdom of Saudi Arabia.

REFERENCES

[1] Appelt, D. and Israel, D., 1999. Introduction to Information Extraction Technology: IJCAI-99 tutorial. [Online]Available at: <http://www.ai.sri.com/~appelt/ie- tutorial/IJCAI99.pdf> [Accessed 13 April 2011].

[2] J Tang, M Hong, D Zhang, B Liang, 2007. Information Extraction: Methodologies and Applications, pp.1-40.

[3] Mooney, R. J., Bunescu. R., 2005.Mining Knowledge from Text Using Information Extraction.ACM SIGKDD Explorations Newsletter. 7(1),pp.3-10.

[4] Zhou, G., Su, J., 2002. Named entity recognition using an HMM-based chunk tagger. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. July 2002 Philadelphia: USA pp.473-480.

[5] Lafferty, J., McCallum, A., Pereira, F.,2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. The 18th International Conference on Machine Learning.

[6] Li, W., McCallum, A., 2003. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons.Proceedings of the 7th conference on natural language learning.

[7] McDonald, R., Pereira, F.,2005. Identifying gene and protein mentions in text using conditional random fields. BMC Bioinformatics. 6(1),pp.1-3.

[8] Christopher J. C. Burges, 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl.Discov., 2(2),pp.121–167.
[9] Pianta, Emanuele, Christian Girardi, and Roberto Zanoli. "The TextPro Tool Suite." LREC. 2008.
[10] Jeremy and JugalKalita, 2009. Extracting Geospatial Entities from Wikipedia. International Conference on Semantic Computing , pages 450 - 457 .

Speech Recognition Techniques on Mobile Devices:- Analysis of Various Approaches

[1]Gulbakshee Dharmale, [2]Dr. Dipti D. Patil, [3]Dr. Vilas Thakare
[1]Research Scholar, 1,3SGB Amravati University, Amravati, INDIA,
[2]MKSSS's CCOEW, SavitribaiPhule, Pune University, Pune, INDIA

ABSTRACT

Speech is the most common and convenient way to communication. Speech is also faster than typing on a keypad and more expressive than clicking on a menu item. For these reason Automatic Speech Recognition (ASR) is become very important and popular in today's world. Speech recognition is the process of converting spoken words into text. After years of research and development the accuracy of automatic speech recognition remains one of the important research challenges (e.g. variations of the context, speakers, and environment).The design of speech recognition system requires careful attentions to different issues such as: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifiers, performance evaluation and database. One of the major problems faced in speech recognition is that the spoken word can be vastly altered by accents, dialects and mannerisms. This paper presents review of Automatic Speech Recognition techniques with Artificial Intelligence related to mobile platform.

Keywords- Automatic Speech Recognition, Speech recognition Techniques, Feature Extraction, Artificial Intelligence.

I. INTRODUCTION

Speech is a tool of communication, also a symbol of identity and authorization. The idea of speech-based recognition comes from the human imagination and creativity that have been frequently used in television programs and several movies. Speech recognition- based authentication has been presented as the symbol of technological advancement as well as a secure system.

Smartphone's and tablets are rapidly overtaking desktop and laptop, computers as people's primary computing device. They are heavily used to access the web, read and write messages, interaction social networks, etc. [1]. This popularity comes despite the fact that it is significantly more difficult to input text on these devices, mostly by using an on- screen keyboard. Automatic speech recognition is alternative to typing on mobile services. It is a natural and increasingly popular. Google offers the ability to search by voice [2] on Android, iOS and Chrome; Apple's iOS devices come with Siri, a conversational assistant. On both Android and iOS devices, users can also speak to fill in any text field where they can type, a capability greatly used to dictate SMS messages and e-mail [3].

Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop systems and techniques for speech input to machine. Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert it to a machinereadable format. In today's world many speech recognition applications, such as voice dialing, simple data entry and speech-to-text are exist. While speech recognition sets its goals at recognizing the spoken words in speech, the main aim of speaker recognition is to identity the speaker by characterization, extraction and recognition of the information confined in the speech signal [4].

As these mobile devices are often used when the person is "on the move", variable acoustic environments and limited resources on the mobile device needs special arrangements. The Automatic Speech Recognition is a software technology that allows a machine to extract the message, oral contained in a speech signal [5]. This technology uses computational methods in areas of signal processing and artificial intelligence.

Fig.1 shows a mathematical representation of speech recognition system in simple equations which contain front end unit, model unit, language model unit, and search unit. The recognition process is shown below (Fig .1).

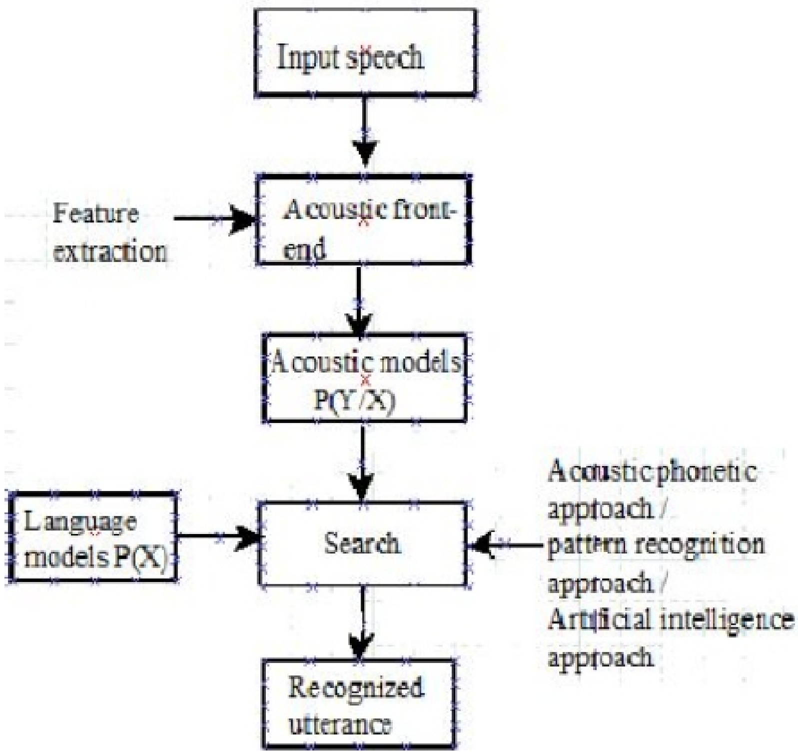


Fig.1. Basic model of speech recognition

The standard approach to large vocabulary continuous speech recognition is to assume a simple probabilistic model of speech production whereby a specified word sequence X, produces an acoustic observation sequence Y, with probability P (X, Y). The goal is then to decode the word string, based on the acoustic observation sequence, so that the decoded string has the maximum a posteriori (MAP) probability.

$$P\left(\frac{X}{Y}\right) = \arg max_X P\left(\frac{X}{Y}\right) \qquad - (1)$$

Using Bay's rule, equation (1) can be written as

$$P\left(\frac{X}{Y}\right) = \frac{P(^{Y}/_{X})P(X)}{P(Y)} \qquad - (2)$$

Since P(Y) is independent of X, the MAP decoding ruleofequation(1) is

$$X = agr\ max_X P\left(\frac{Y}{X}\right) P(X) \qquad - (3)$$

The first term in equation (3) P (Y/X), is generally called the acoustic model, as it estimates the probability of a sequence of acoustic observations, conditioned on the word string. Hence P(Y/X) is computed. It is necessary to build statistical models for sub word speech units, build up word models from these sub word speech unit models (using a lexicon to describe the composition of words) and then postulate word sequences and evaluate the acoustic model probabilities via standard concatenation methods, for large vocabulary speech recognition systems. The second term in equation (3) P(X), is called the language model. It describes the probability associated with a postulated sequence of words. Such language models can combine both syntactic and semantic constraints of the language and recognition task.

II.RELATED WORK

Gaussian mixture models (GMMs) is used by most conventional speaker recognition systems to capture framelevel characteristics of a person's voice, where the speech frames are assumed to be independent of one another as the physical shape of a human vocal tract is different from person by person. Hence, each human speaks in a different way. If a person is asked to utter the same word twice, the speech signal will not be exactly same as the frequency and other sound properties may differ from time to time. There are some features that makes difficult to understand speech signals these are; an environment where human speaks, the dialect of the language, differences in the vocal tract length of males, female and children provide the speech variation. Though, there are still some features in the human speech which can be mathematically modelled and used for predicting words from it but it demands tremendous amount of time and effort [6]. To model a human hearing system, it is important to understand the working of human auditory system which is shown in table 1 given below:

Table 1: Working of human auditory system

Level	Action performed
The linguistic level of communication [6]	1. The idea is formed in the mind of the speaker. 2. The idea is then transformed to words, phrases and sentences according to the grammatical rules of the language
The physiological level of communication [6]	1. The brain creates electric signals that move along the motor nerves. 2. Then these electric signals activate muscles in the vocal track and vocal cords.

Because of this independence assumption, GMMs often fail to capture certain types of speaker-specific information that evolve over time scales of more than one frame. For example, since words usually span many frames, GMMs [7] tend to be poorly suited for modelling differences in word usage (idiolect) between speakers. In recent times, automatic speaker recognition research has expanded from utilizing only the acoustic content of speech to examining the use of higher levels of speech information, commonly referred to as "highlevel features." Anencouraging direction in high-level feature research has been the use of n-gram based models to capture speaker specific patterns in the phonetic and lexical content of speech.

In, Doddington performed an important study about using the lexical content of speech for speaker recognition and an ngram based technique is introduced for modelling a speaker's idiolect. This trend in research was continued by Andrews, Kohler, and Campbell among others, who used similar n gram, based models to capture speaker pronunciation idiosyncrasies through analysis of automatically recognized phonetic events. This line of research is generally referred to as "Phonetic Speaker Recognition." The research of Andrews et al. and Doddington showed word and phone n-gram based models to be quite promising for speaker recognition.

There have been numerous attempts, especially since the Johns Hopkins 2002 Workshop to harness the power of all kinds of high-level features. The relative frequencies of phone n-grams as features for training speaker models and for scoring test-target pairs used by the current "state-of-the-art" in phonetic speaker recognition[8]. Typically, these relative frequencies are computed from a simple 1-best phone decoding of the input speech. The phonetic speaker recognition research work has been extended in various ways by introducing different modelling strategies and different methods of utilizing the source information such as described in Navratil. It proposed a method involving binary-treestructured statistical models for extending the phonetic context beyond that of standard n-gram (particularly bigrams) by exploiting statistical dependencies within a longer sequence window without exponentially increasing the model complexity, as is the case with n-grams. The described approach confirms the relevance of long phonetic context in phonetic speaker recognition and represents an intermediate stage between short phone context and word-level modeling without the need for any lexical knowledge. Binary-tree model represents a step towards flexible context structuring and extension in phonetic speaker recognition, consistently outperforming standard smoothed bigrams as well as trigrams.

A conditional pronunciation modelling method is proposed by Klusacek. It uses time-aligned streams of phones and phonemes to model a speaker's specific pronunciation. The system uses phonemes drawn from a lexicon of pronunciations of words recognized by an automatic speech recognition system to generate the phoneme stream and an open-loop phone recognizer to generate a phone stream. The phoneme and phone streams are aligned at the frame leveland conditional probabilities of a phone, given phoneme, are estimated using co-occurrence counts. A probability detector is then applied to these probabilities for the speaker detection task. This approach achieves a relatively high accuracy in comparison with other phonetic methods in the Super SID project at the Johns Hopkins 2002 Workshop.

III. EFFECTIVE TECHNIQUES FOR AUTOMATIC SPEECH RECOGNITION

Basically there exist three techniques to speech recognition.

These are:

1. Acoustic Phonetic Approach [9]
2. Pattern Recognition Approach [10]
3. Artificial Intelligence Approach [5, 9] Classification of these speech recognition techniques shown in given tree diagram.

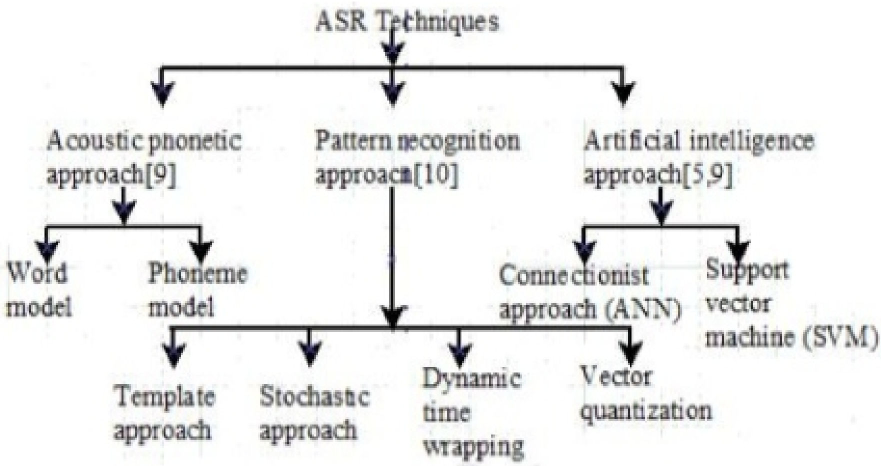


Fig. 2. Automatic Speech Recognition Approaches

3.1 Acoustic Phonetic Approach

Acoustic phonetic approach is the oldest approach to speech recognition were based on finding speech sounds and allocate categorized labels to these sounds which is the basis of the acoustic phonetic approach, which postulates that there exist finite, distinctive phonetic units in spoken language and these units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time. There are three basic steps for the acoustic phonetic approach. These are as follows [9]:

The first step is a spectral analysis which combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units.

The second step is speech segmentation and labeling phase in this phase the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels to each segmented region, resulting in a phoneme lattice characterization of the speech.

The third and last step in this approach is validation phase which attempts to determine a valid word (or string of words) from the phonetic label sequences produced by the segmentation to labelling. In the validation process, linguistic constraints on the task (i.e., the vocabulary, the syntax, and other semantic rules) are invoked in order to access the lexicon for word decoding based on the phoneme lattice.

There are two types of acoustic models i.e. word model and phoneme model. An acoustic model is implemented using different approaches such as HMM, ANNs, Dynamic Bayesian Networks (DBN), Support Vector Machines (SVM). HMM is used in some form or the other in every state of the art speech and speech recognition system [9].

3.2 Pattern Recognition Approach

Pattern training and pattern comparison are involves in the pattern-matching approach. The fundamental feature of pattern-matching approach is that it uses a well formulated mathematical framework and creates consistent speech pattern representations for reliable pattern comparison from a set of labeled training samples via a formal training algorithm. A speech pattern representation can be in the form of a speech template or a statistical model and can be applied to a sound (smaller than a word), a word ora phrase. In the pattern-comparison stage of the approach, a direct comparison is made between the unknown speeches (the speech which needs to be recognized) with each possible pattern learned in the training stage in order to determine the identity of the unknown according to the goodness of match of the patterns. Usually, pattern recognition approaches are model based such as Hidden Markov Model (HMM), Artificial Neural Networks (ANN), Support Vector Machine (SVM), Vector Quantization (VQ) and Dynamic Time Warping (DTW) [10].

3.2.1 Template Approach

A collection of prototypical speech patterns are stored as reference patterns representing the dictionary of candidate's words. Recognition is then carried out by matching an unknown spoken utterance with each of these references.

There are two main ideas in template method these are:

1. One key idea is to derive typical sequences of speech frames for a pattern (a word) via some averaging procedure and to depend on the use of local spectral distance measures to compare patterns.
2. Another key idea is to use some form of dynamic programming to temporarily align patterns to account for differences in speaking rates across speakers as well as across repetitions of the word by the same speaker [9].

3.2.2 Stochastic Approach

Stochastic modelling entails the use of probabilistic models to deal with incomplete or uncertain information. In speech recognition, uncertainty and incompleteness arise from many sources; for example, contextual effects, speaker variability, confusable sounds and homophones words [9].

A common issue in pattern recognition is high dimensionality of feature vectors. There are many reasons for having high dimensional feature spaces. For instance, static features can be augmented with dynamic spectral information in speech feature extraction. One way of achieving this is by combining multiple consecutive Mel-filtered cepstrum coefficients (MFCC) feature vectors to form high dimensional super vectors that may represent on the order of 100 milliseconds of speech. These super feature vectors can have very high dimensionality, which may lead to significant problems when performing a pattern recognition task. Therefore, it is a good practice to perform some sort of dimensionality reduction before applying a particular pattern recognition algorithm to these features. Intuitively, a good dimensionality reduction algorithm should be able to preserve important information from the original feature space in the low dimensional transformed feature vectors [11].

3.2.3 Dynamic Time Warping (DTW)

Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW is a well-known application of automatic speech recognition used to deal with different speaking speeds. In general, DTW is a method that allows a computer to find an optimal match between two given sequences (e.g., time series) with certain restrictions. That is, the sequences are "warped" non-linearly to match each other. This sequence alignment method is often used in the context of hidden Markov models [12].

3.2.4 Vector Quantization (VQ)

The objective of VQ is the formation of clusters, each representing a specific class. During the training process, extracted feature vectors from each specific class are used to form a code book, through the use of an iterative method. Thus, the resulting code book is a collection of possible feature vector representations for each class. During the recognition process, the VQ algorithm will go through the whole code book in order to find the corresponding vector, which best represents the input feature vector, according to a predefined distancemeasure[13].

3.3 Artificial Intelligence Approach (Knowledge Based Approach)

Artificial intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach. In this, it exploits the concepts and ideas of acoustic phonetic and pattern recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram [10]. Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based approaches have been very effective in the design of a variety of speech recognition systems; it provided little insight about human speech processing, thereby making error analysis and knowledge-based system enhancement difficult [9]. There are different techniques and algorithms comes under Artificial Intelligence approach.

3.3.1 Connectionist Approaches (Artificial Neural Networks)

The artificial intelligence approach try to automate the recognition procedure according to the way a person applies intelligence in visualizing, analyzing, and characterizing speech based on a set of measured acoustic features. Among the techniques used within this class of methods are uses of an expert system (e.g., a neural network) that integrates phonemic, lexical, syntactic, semantic and even pragmatic knowledge for segmentation and labeling that uses tools such as artificial Neural Networks for learning the relationships among phonetic events. The focus in this approach has been generally in the representation of knowledge and integration of knowledge sources.

Connectionist modeling of speech is the earliest development in speech recognition and still the subject of much controversy. In connectionist models, knowledge or constraints are distributed across many simple computing units instead of encoded in

individual units, rules or procedures. Uncertainty is modeled not as likelihoods or probability density functions of a single unit, but by the pattern of activity in many units. The computing units are simple in nature and knowledge is not programmed into any individual unit function; rather, it lies in the connections and interactions between linked processing elements. The uniformity and simplicity of the underlying processing element makes connectionist models attractive for hardware implementation, which enables the operation of a remaining to be simulated efficiently [9].

3.3.2 Support Vector Machine (SVM)

SVM is one of the most efficient machine learning algorithms, which is mostly used for pattern recognition. The basic idea of the SVMs is building an optimal hyper plane in order to use in classification of linearly separable patterns. An optimal hyper plane is a selected one from a set of hyper planes which maximizes hyper plane margin, which is the distance from hyper plane to the nearest point of the pattern. SVM is primarily set to maximize the margin, which will guaranty that the input pattern would be classified correctly. In order to classify data, based on either a priori information or statistical data mined from raw data set, pattern recognition is widely utilized and this makes it an extremely powerful tool for data separation in many fields. The support vector machine (SVM) usually copes with pattern classification that means this algorithm is used mostly for classifying the different types of patterns. Now, there are two different type of patterns such as Linear and non-linear. Linear patterns areeasily distinguishable or can be easily separated in low dimension whereas non-linear patterns are not easily distinguishable or cannot be easily separated and hence these type of patterns need to be further manipulated so that it can be easily separated [14].

IV. DIFFERENT SPEECH RECOGNITION BLOCKS

A general speech recognition system consists of four blocks:

- Feature extraction, Language modelling, Pronunciation modelling, Decoder. This blocks help to simplify the face recognition models.

4.1 Feature Extraction

In speech recognition, feature extraction requires much attention because recognition performance depends heavily on this phase. Passing huge speech data as inputs to a machine learning algorithm is an exhaustive task with no extra merits for the size of data. Large data set might be an obstacle for enrollment and training stages which may prevent the system from learning. It greatly affects the system performance. Due to that, the input data set should be transformed into a reduced representation set of features this transformation process is called feature extraction. The features should be representative and carefully selected which leads to better and faster recognition [7]. The process of feature extraction reduces speech data while maintaining the discriminative information of a speaker voice signal [10]. There are different techniques available for the feature extraction of speech like MFCC, RAST, PLP, LPCC, but MFCC is the most commonly used feature extraction technique for speech[5]. Table 2 shows the advantages and disadvantages of different features extraction techniques used for the voice recognition.

Table 2.Advantages and disadvantages of feature extraction techniques [13].

Feature Extraction Technique	Advantages	Disadvantages
MFCC [13][15]	1. MFCC Provides good discrimination. 2. Low correlation between coefficients 3. MFCC is similar to the human auditory perception system; as itsnot based on linear characteristics. 4. Important phonetic characteristics can be captured by MFCC. 5. MFCC analysis lookalikes the behavior of human auditory system which responds linearly to the low frequency and logarithmic for high frequency.	1. Low robustness to noise. 2. In a continuous speech environment, a frame may not contain information of only one phoneme, but of two consecutive phonemes. 3. Only the power spectrum is considered, ignoring the phase spectrum of speech signals hence Limited representation of speech signals.

DWT [13][15]	1. DWT also considers temporal information present inspeech signals, apart from the frequencyinformation. 2. DWT is able to perform efficient time and Frequency Localisations. 3. Successfully used for de-noising tasks. 4. Capable of compressing a signal without major Degradation. 5. DWT calculates an optimal warping path between two time series of different lengths.	1. DWT is not flexible since the same basic wavelets have to be used for all speech signals
WPT[13]	1. Same as DWT, but WPT shows also further detail present in the high frequency bands.	1. Not flexible since the same basic wavelets have to be used for all speech signals
LPC[3],[15][16]	1. Spectral envelope is represented with lowdimension feature vectors. 2. LPC method is simple to implement andmathematically precise. 3. LPC analysis provides better representation as it closely matches the resonant structure of human vocal tract that produces the corresponding sound. 4. The key idea behind linear prediction is to extract the vocal tract parameters	1. Linear scales are inadequate for the representation of speech production or perception. 2. Feature components are highly correlated. 3. LPC cannot include a priori information on the speech signal under test. 4. The Linear Prediction (LP) models the input signal with constant weighting for the whole frequency range.
PLP [7][13]	1. Reduction in the discrepancy between voiced and unvoiced speech.0 2. PLP discardsirrelevant information of the speech based on the concept of psychophysics and thus improves recognition. 3. Resultant feature vector is low-dimensional. 4. PLP is based on short term spectrum of the speech Signals	1. Resultant feature vectors are dependent on the whole spectral balance of the formant amplitudes. 2. Spectral balance is easily altered by the communication channel, noise, and the equipment used.
RASTA-PLP[13]	1. spectral components that change slower orquicker than the rate of change of the speech signal are suppressed. 2. the RASTA-PLP outperformed PLP, by obtaining an increase in the accuracy.	1. Poor performance in clean speech environments.

VQ[13]	1. Reduction in the required memory storage size for the spectral analysis information. 2. Reduction in the computational cost for thecalculation of similarity between feature vectors. 3. Discrete representation of speech signal. 4. Fast training speed.	1. Training time increases linearly with increase in vocabulary size. 2. Quantisation error in the discrete representation of speech signals. 3. Temporal information is ignored.

4.2 Decoder

Decoding is the most essential step in the speech recognition process. Decoding is performed for finding the best match for the incoming feature vectors using the knowledge base [5]. The objective of the decoder is to find out the most probable sequence of words from the language model, produces the observation sequence[17].

The Viterbi beam search algorithm is used by the decoding stage to find the most likely sequence of phones for the observed speech. Each stage in the algorithm uses models, which represent the probabilities of sounds, sequences of sounds, words and sequences of words in the language. Gaussian distributions are used to represent nature of the sounds and HMMs are used to model sequences and duration of the sounds. Word sequences and their probabilities are stored as weights, which are added during the decode process [17].

4.3 Language Modeling

Language models are used to guide the search correct word sequence by predicting the possibility of nth word using (n-1) preceding words. Language models can be classified into: [5]

1. Uniform model: In uniform model each word has equal probability of occurrence.
2. Stochastic model: In stochastic modelprobability of occurrence of a word depends on the word preceding it.
3. Finite state languages: languages use a finite state network to define the allowed word sequences.
4. Context free grammar: It can be used to encode which kind of sentences is allowed.

4.4 Pronunciation Modelling

In pronunciation modelling, during recognition the sequence of symbols generated by acoustic model HMM is compared with the set of words present in dictionary to produce sequence of words which is the system's final output contains information about which words are known to the system and how these words are pronounced i.e. what is their phonetic representation. Decoder is then used for recognizing words by combining and optimizing the information of acoustic & language models [5].

V. ADVANCED ASR BASED TECHNIQUES

There are different basic ASR techniques which are modified by different researchers for the speech recognition in the mobile devices is called advanced ASR based techniques. Some of the sear studied in this section.

5.1 Large Vocabulary Continuous Speech Recognition

Research and development activities in the area of Large Vocabulary Continuous Speech Recognition (LVCSR) are concentrated on developing a dictation system. There are also noise models trained for better modelling of the inter word and inter sentence noise which could produce false tri-phones detection. In the language modelling task during the last years, it has encountered several problems with text pre- processing, selection of the basic statistical methods used in the modeling of the other languages and adaptation into the area of application. Another important part in the process has been optimization of the resultant model, which introduced phonetic and linguistic relations between words. These optimization steps have caused an improvement in recognition accuracy of the LVCSR system [18].

5.2 An Arabic Speaker Verification System

In research work, it developed and analysed an Arabic speaker verification system with good accuracy. The system is used in an access control application for mobile devices to prevent unauthorized users from gaining access to the device. The process of speaker verification consists of two main stages; enrolment stage and verification stage [19].

Enrolment stage: In the enrolment stage, a speaker S repeats a set pass phrase n times. The speech signals of the speaker are pre-processed and features are extracted. After that, the features are passed to the support vector machine (SVM) as the enrolment data. By training SVM, a speaker statistical model for S is created. A background model for imposters is created as well; which uses speech signals from a variety of unauthorized speakers.

Verification stage: During the verification stage, a speaker X says the pass phrase and claims to be speaker S. The claimed speaker model and background model are fetched to verify the claimed identity. Similarly, the spoken phase goes through the same phases in the enrolment stage.

5.3 Android Applications using Voice Controlled Commands

Assistive Technology (AT) is dedicated to increasing the independence and mobility of the persons with disabilities. Persons with quadriplegia are affected by limitations in physical and independence. In this project, voice recognition is explored as a template upon which the independence of persons with neuromuscular disorders can be expanded. For this reason, android applications were developed on a Smartphone to operate a television remote via Bluetooth exchange and PIC processing.

The hardware implements a Bluetooth modem, BlueSMiRF silver (WRL-10269), which receives the commands from the android application software that was developed for the Smartphone. The BlueSMiRF modem communicates with the PIC ®Microcontroller (PIC18f4525- I/P) via the EUSART transmit and receive pins on the microcontroller, pins 25 and 26, respectively. Each function in the android application recognizes keywords which then send a specific signal via Bluetooth connection to the microcontroller. A simple code in the microcontroller interprets the signals received from the modem and sends a corresponding signal out of one of the seven outputs of port B to the quad bilateral switches. In order to trigger specific buttons on the direct TV remote control, reverse engineering was used by soldering wires from a ribbon cable to the specific button locations inside the remote [20].

5.4 Augmented Reality

Augmented reality enhances user perception by supplementing the real world through the additional of virtual content which is stored in a library. Augmented reality enables users to complement their actual environment with that of a simulated environment to provide enhanced information and data without sacrificing the information stored within the real environment. Augmented reality is used to describe a system that superimposes computer generated information overlaying the real environment. The goal of augmented reality is to superimpose information in the form of audio, text, graphics and other sense enhancements over a real environment in real time.

Upon viewing potential mobile devices to serve as training tools, augmented reality is selected as a potential mobile technology. Augmented reality is selected here because it offers hands free operation through a head mounted display HMD. System is capable of transferring data between the associate and system without hand operation or relocation to a computer. User is able to control the device through audible commands and incorporates a simple and easy to use operator interface. The key benefit of this implementation is it will minimize, or potentially eliminate, the need for a training associate. A cost savings could be directly associated with implementation as a trainer will not be needed for extended periods of time [21].

VI. ARTIFICIAL INTELLIGENCE BASED SPEECH RECOGNITION

The artificial intelligence approach attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analyzing and finally making a decision on the measured acoustic features. Expert system is usedwidely in this approach (Mori et al., 1987) [22].

The Artificial Intelligence approach [23] is a fusion of the acoustic phonetic approach and pattern recognition approach. In this, it exploits the ideas and concepts of Acoustic phonetic and pattern recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based approaches have been very effective in the design of a variety of speech recognition systems; It provided little insight about human speech processing, thus making error analysis and knowledge based system enhancement difficult. On the other hand, a large body of linguistic and phonetic literature provided insights and understanding to human speech processing [24].

In its pure form, knowledge engineering design involves the direct and explicit in corporation of expert's speech knowledge into a recognition system. This knowledge is usually derived from careful study of spectrograms and is incorporated using rules or procedures. Pure knowledge engineering was also motivated by the interest and research in expert systems. However, this approach had only limited success, largely due to the difficulty in quantifying expert knowledge. Another difficult problem is the integration of many levels of human knowledge phonetics, lexical access, syntax, semantics and pragmatics. Alternatively, combining independent and asynchronous knowledge sources optimally remains an unsolved problem. In more indirect forms, knowledge has also been used to guide the design of the models and algorithms of other techniques such as template matching and stochastic modeling. This form of knowledge application makes an important distinction between knowledge and algorithms enable us to solve problems. Knowledge enables the algorithms to work better. This form of knowledge based system enhancement has contributed considerably to the design of all successful strategies reported. It plays an important role in the selection of a suitable input representation, the definition of units of speech, or the design of the recognition algorithm itself.

CONCLUSION

Speech is the most convenient means of communication between people. Because of the technological curiosity to build machines that mimic humans or desire to automate work with machines like mobile phones, research in speech and speaker recognition, as a first step toward natural human- machine communication, has attracted much enthusiasm over the past five decades. In this paper the detail review on Speech recognition with Artificial Intelligence related to mobile device is done. The main aim of this paper is to study the Speech recognition techniques and how the artificial intelligence can be used for speech recognition purpose.

REFERENCES

[1] Anuj Kumar, AnujTewari, Seth Horrigan, Matthew Kam1, Florian Metze and John Canny, "Rethinking Speech Recognition on Mobile Devices", IUI4DR, California, USA, February 13, 2011, pp.1-6.

[2] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope, "Google search by voice: A case study in Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics", Springer, 2010, pp. 61–90.

[3] B.Ballinger,C. Allauzen, A. Gruenstein, and J. Schalkwyk, "Ondemand language model interpolation for mobile speech input," in Proc. Interspeech, 2010.

[4] SadaokiFurui, "50 years of Progress in speech and Speaker Recognition Research", ECTI Transactions on Computer and Information Technology,Vol.1. No.2, November 2005.

[5] PreetiSaini, ParneetKaur,"Automatic Speech Recognition: A Review", international Journal of Engineering Trends and Technology- Volume4, Issue2, ISSN: 2231-5381, 2013,pp.132-136.

[6] Shanthi Therese S., Chelpa Lingam," Review of Feature Extraction Techniques in Automatic Speech Recognition", International Journal of Scientific Engineering and Technology (ISSN: 2277-1581), Volume No.2, Issue No.6,1 June 2013, IJSET@2013,pp : 479-484.

[7] AbdulrahmanAlarifi, IssaAlkurtass," SVM based Arabic Speaker Verification System for Mobile Devices", International Conference on Information Technology and e-Services, 978-1-4673-1166- 3/12/IEEE ©2012 .

[8] Nicholas Mulhern, Neil McCaffrey, Nicholas Beretta, Eugene Chabot, Ying Sun, "Designing Android Applications using Voice Controlled Commands For Hands free interaction with Common Household Devices", 39th Annual Northeast Bioengineering Conference, IEEE 2013, pp. 265-266.

[9] Sanjivani S. Bhabad ,Gajanan K. Kharate, "An Overview of Technical Progress in Speech Recognition", International Journal of Advanced Research in Computer Science and Software EngineeringResearch Department of E & TC, Pune university, India, Volume 3, Issue 3, March 2013, pp. 488-497.

[10] ShufeiDuan • Jinglan Zhang • Paul Roe • Michael Towsey, "A survey of tagging techniques for music, speechand environmental sound",ArtifIntell Rev DOI 10.1007/s10462-012-9362-y © Springer Science+Business Media Dordrecht 2012.

[11] Vikrant Singh Tomar, "A Family of Discriminative Manifold Learning Algorithms and Their Application To Speech Recognition", IEEE/ACM transactions on audio, speech, and language processing, vol. 22, no. 1, January 2014,pp.161-171.

[12] Parwinder pal Singh Er. Bhupindersingh, "Speech Recognition as Emerging Revolutionary Technology" , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012.

[13] Michelle Cutajar, Edward Gatt, Ivan Grech, Owen Casha, Joseph Micallef, "Comparative study of automatic speech recognition techniques",IET Signal Process, Vol. 7, Issue 1, 2013, pp. 25–46.

[14] SasanKaramizadeh, Shahidan M. Abdullah , MehranHalimi , JafarShayan and Mohammad javadRajabi, "Advantage and Drawback of Support Vector Machine Functionality", International Conference on Computer, Communication, and Control Technology (I4CT 2014), September 2 - 4, 2014 - Langkawi, Kedah, Malaysia.

[15] Sharada C. Sajjan, Vijaya C, "Comparison of DTW and HMM for Isolated WordRecognition", International Conference on Pattern Recognition, Informatics and Medical Engineering, March 21-23, 978-1-4673-1039-0/12/$31.00 ©2012 IEEE

[16] Nagsen S. Bansod1, Siddharth B. Dadhade2, "Speaker Recognition using Marathi (Varhadi) Language", 978- 1-4799-3966- 4/14 $31.00 © 2014 IEEE DOI 10.1109/ICICA.2014.

[17] Ojas A. Bapat, P. D. Franzon. R. M. Fastow, "A Generic and Scalable Architecture for a Large Acoustic Model and Large Vocabulary Speech Recognition Accelerator Using Logic on Memory", IEEE transactions on very large scale integration (VLSI) systems, 2014,pp.1-12 .

[18] Deng and D. Yu, "Use of differential cepstra as acoustic features in hidden trajectory modelling for phonetic recognition", in Proc.ICASSP, 2007,pp. 445–448.

[19] Bilmes, J. ; Das, S. ; Duta, N, "Novel approaches to Arabic speech recognition: report from the 2002 Johns-Hopkins Summer Workshop" International Conference on Acoustics, Speech, and Signal Processing Proceedings. (ICASSP '03) IEEE 2003.

[20] C.H.Lee, etc.al., "Acoustic modeling for large vocabulary speech recognition",Computer Speech and Language 4, 1990,pp.127- 165.

[21] MatusPleva, StanislavOndas, JozefJuhar, Anton Cizmar, Jan Papaj and LubomirDobos, "Speech and Mobile Technologies for Cognitive Communication and Information Systems",2011.

[22] BeshoyMorkos, TorstenRilka, " Mobile devices within manufacturing environments: a BMW applicability study ",Int J Interact Des (2012),10 April 2012 © Springer-Verlag2012,pp:101– 111.

[23] J.Ferguson, Ed., "Hidden Markov Models for Speech", IDA,Princeton, NJ,1980.

[24] B.H.Juang and S.Furui,"Automatic speech recognition and understanding: A first step toward natural human machine communication", Proc.IEEE, 88, 8, 2000, pp.1142-1165.

Scorm Based Solution for LMS Integration and Performance Enhancement

[1]Boby Antony, [2]J. Meenakumari
[1]Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India
[2]Professor & Dean, Department of Computer Science and Applications,
The Oxford College of Science, Bangalore

ABSTRACT

Transformation has been happening in every field with the help of Information and Communication Technology (ICT). ICT is revolutionizing every aspect of an education sector in this digital era. To enhance the sustainability of higher education it is very much required to integrate ICT-based tools. This study analyses the key features of ICT-based Learning Management Systems (LMS) so as to enhance the efficiency of teaching-learning process in higher education. This paper also proposes a SCORM enabled technological hybrid LMS framework as a solution for the challenges faced by present LMS.

Keywords— SCORM, ICT, digital era, sustainability, LMS, teaching-learning

I. INTRODUCTION

Information and Communication Technologies (ICT) has penetrated into all walks of life and education sector is no exception. Many studies reveal that ICT is an extremely powerful tool that could bring about tremendous changes in the education process. Innovation has become mandatory to sustain in the global competitive environment. ICT plays an important role in reaching various goals.

Its ability to transcend time and space allows learning to take place 24 hours a day, 7 days a week eliminating geographical barriers [1]. "Technology is not a panacea for educational reform, but it can be a significant catalyst for change. To those looking for a simple innovative solution, technology is not the answer. To those looking for a powerful tool to support collaborative learning environments, technology holds tremendous potential"[].

LMS Definitions and Features

The Learning Management System (LMS) is an important tool for the development of curriculum design, management of students' learning and their motivation to learn. The LMS is also useful in the development of student assessment. The LMS can manage all teaching and learning processes of registration, scheduling, checking availability of content, tracking the performance of the learner and issuing reports about it, facilitating communication among teachers and learners, etc. [3].

Many other terms also exist with LMS with similar meaning, such as course management system, content management system and e-learning platform. Alias and Zainuddin defined learning management system as “a software application or web-based technology used to plan, implement, and assess a specific learning process” [4]. Baumgartner and Maier define LMS as “a server-side installed software, which assists in teaching of any learning material via the internet and supports the organisation of the necessary processes” [5]. “LMS means a suite of functionalities designed to deliver, track, report and manage learning content, student progress and student interactions. The term ‘LMS’ can apply to very simple course management systems, or to highly complex enterprise-wide distributed environments” [6]. Learning management systems such as Moodle, Blackboard, WebCT etc. focus on supporting teachers and administrators in creating, administering, and managing online courses. LMSs provide a great variety of features which can be included in the courses. They have become very successful in technology enhanced learning and are commonly used by educational institutions.

Learning management systems are of two types: proprietary systems and open-source systems. Blackboard, Desire2Learn and eCollege are examples for proprietary systems whereas moodle and sakai are the best examples for the open-source systems. Each of the learning management system has got innumerable features such as: grades, calendar, wiki, lessons and workshop, chat and messages, forum, notes and resources, online examinations, assessment, quizzes, uploading and sharing notes, reporting etc. [7]. The initial versions of an LMS focused on organizing and managing course content and learners. In the initial stage all the LMS were stand-alone systems which support the educational and administrative needs of the institution.

The Introduction of Web 2.0 technologies brought about tremendous changes in the Learning Management Systems. It facilitates the usage of wikis, blogs, RSS, 3D virtual learning etc. It has made substantial changes in the student attitude from passive to active learners. LMS gives access to synchronous as well as asynchronous learning resources and activities. The current LMSs are hosted in the cloud, which makes the institutions and companies free from installing and maintaining in- house systems. Many of the current LMS provides features like mobile learning, virtual class room, video conferencing etc.

The ability to access the instructional components and contents from any location makes LMS user friendly and convenient. The new generation LMS reduces the time and cost in delivering the instructional content and make it affordable. They are capable of accommodating frequent changes happening in the technology changes without reworking the entire system. It has the ability to use instructional components in various applications [8].

A detailed study was conducted on the leading LMS available today. The important features and parameters of the existing LMS are identified. It was found that the current LMS are not capable of meeting all the requirements of current education scenario. SCORM (Sharable Content Object Reference Model) enabled LMS is found as a solution to meet the challenges. A hybrid SCORM enabled LMS model proposed here is capable of meeting the challenges and provides a better teaching-learning experience.

II. THEORETICAL BACKGROUND OF SCORM

SCORM (Sharable Content Object Reference Model) is a package of technical standards developed by the ADL (Advanced Distributed Learning), a research wing of Department of Defence in US. This relates to the development of common specifications and standards for technology-based learning deployed over the internet. These standards enable consistency in finding, importing, sharing, reusing, and exporting learning content in web-based learning and content management systems. It also facilitates user tracking and generation of reports based on learning objectives. SCORM enabled standardization of the method of communication between eLearning courses and SCORM conformant learning and content management systems. SCORM presently includes the following elements [9]:

- An Application Programming Interface (API) to communicate information on interaction by learners with content objects.
- A well-defined data model to represent information regarding user interaction with content objects.
- A content packaging specification that enables interoperability of learning content
- A standard set of metadata elements that can be utilized to describe learning content and a set of standard sequencing rules that could be applied to the organization of learning content SCORM is a set of technical standards for e-learning software products. SCORM facilitates programmers to write code so that it can be plugged in with other e- learning software. It is the de facto industry standard for interoperability in e-learning. SCORM governs the aspect of communication between online learning content and Learning Management Systems (LMS). SCORM is purely a technical standard [10].

The SCORM package consists of three parts:

- Run-time Environment (RTE)
- Content Aggregation model (CAM)
- Sequencing and Navigation (SN)

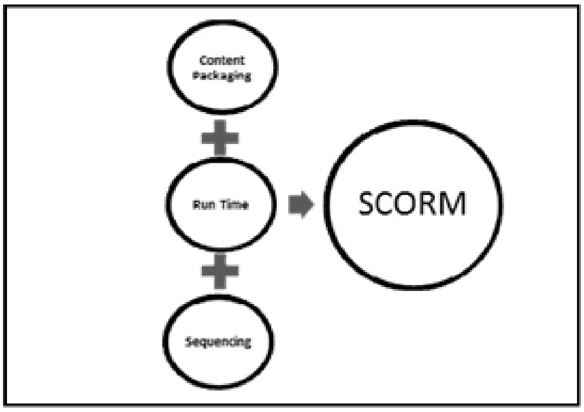


Figure 1: SCORM Package

The CAM [11] comprises regulations for identifying components used in a learning experience, and supports storage, labelling, packaging, exchanging and discovery of content. This model is based on the IMS Content Packaging Information Model [12]. A SCORM Content Package normally represents a course, lesson and module. It can also be a group of related content objects. There are three components such as Assets, Sharable Content Objects (SCOs) and Content Organizations. These components are enabled by the RTE (Run-Time Environment).

A. The Run-Time Environment (RTE)

The SCORM RTE loads the entire content to a web browser. LMS initiates this process and it requires an API for implementing this task. This API comprises eight functions using which it accesses the content and communicates with the LMS. The eight base functions are given below:

- Initialize()
- Terminate()
- GetValue()
- SetValue()
- Commit()
- GetLastError()
- GetErrorString()
- GetDiagnostic()

This API adapter applied by the SCORM is a java Script object. The API adapter resides in the parent frame of the window which contains the content. The LMS can launch the content in a new window or in a frameset. The API adapter implements the above mentioned eight base functions to complete the task. Every communication that takes place between the content and the LMS are controlled by the adapter. It reduces the overhead of the content author from searching and communicating with the server instead it only needs to find the API adapter and make the corresponding JavaScript calls. This separation between the server and the client make the portability an easy task. The content can communicate with the LMS only through this JavaScript API Adapter and no other way for the content to communicate with the LMS through any other methods like web services or HTTP requests.

The minimum requirement for a SCORM operation is the call of two basic functions such as the initialize() and Terminate() to start and terminate an operation. All the other functions are used in advance operations in the real time situations for higher level of interaction such as report test results, track time, bookmarking etc. The GetValue() and SetValue() functions are used for read and write operations using data models. The Commit() function is called after any values are set to ensure that the data is persisted. The last three functions are used for the error handling mechanism.

B. The Content Aggregation Model

The Content Aggregation Model consists of three segments:

- Content Model
- Metadata
- Content Packaging

The content model handles the details of the content delivered. It describes the relationship among various modules in the content which is also known as content aggregation. The aggregation is designed in a tree structure. It also describes the physical structure of the files or the content. This model breaks the content into arbitrarily sized units which enables reusability an easy task. These basic units are known as SCOs (sharable content objects) and assets (the basic units). An Asset is an "electronic representation of media, text, images, sound, web pages, assessment objects or other pieces of data". Examples of Assets include images, sound clips, Flash movies, etc. A SCO is a collection of one or more assets that represents a logical unit of learning [13].

The Metadata section offers tools to define the content by means of a pre-defined and public vocabulary. The associated metadata describes any part of the SCORM.

The Content Packaging section describes the ways and means by which the Content Model and Metadata are executed. An XML format data structure is used in both the sections. An XML file named "imsmanifest.xml" contains the details of the entire course. The entire content is packaged in a similar manner to make the system interoperable and reusable. The Content Packaging specification needs all content to be moved in a folder or a ZIP file termed as a "package interchange file (PIF)". The "imsmanifest.xml" must be placed at the root of the PIF

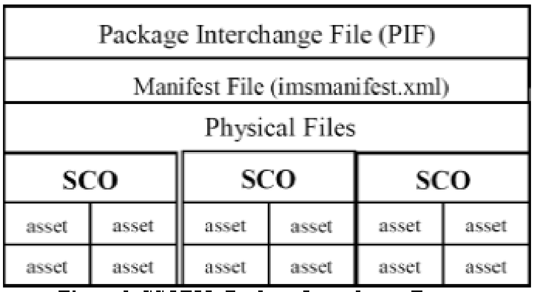


Figure 2: SCORM: Package Interchange Format

C. The Sequencing and Navigation Specification

The Sequencing and Navigation specification manages the navigation between SCOs. An SCO which says nothing about the ways to handle the navigation inside of an SCO is not at all applicable to SCORM courses. It is totally left to the decision of the content developer. The sequencing specification is essential to permit developers of content chunked into SCOs to govern the user experience. In the process of sequencing, the aggregations and SCOs are referred to by the generic term "activities". Content authors specify sequencing rules through XML in the course's manifest.

Sequencing functions on a "tracking model" which strictly parallels a subset of the data model elements exchanged between the content and the LMS at run- time. When a SCO exits, the run-time data for the current SCO is transferred over to the sequencing tracking data. The LMS then invokes the "sequencing loop" which is a set of well-defined algorithms that apply the sequencing rules to the current set of tracking data to determine which activity should be delivered next" [14].

III. HYBRID TECHNOLOGY FRAMEWORK

SCORM enabled hybrid technology framework has been designed to overcome the limitations of existing LMS models. The framework is designed at both macro and micro levels addressing the LMS gap identified with the above concepts as a base Sampling Procedure

A. LMS Technology Framework: Macro Level

The macro framework is designed such that it establishes a communication between the client side content and host system called the run-time environment, which is supported by an LMS. Here the SCORM package defines and controls how teaching content may be packaged into a transferable ZIP file called 'Package Interchange Format' (PIF). Figure 3 shows the macro level LMS technology framework.

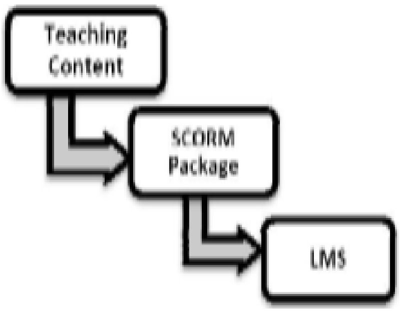


Figure 3: LMS Macro Framework

The teaching content may be of different formats such as lessons or courses with text, image, audio, video etc. and ilable as an online eLearning content. It is also required for the instructor to track the progress of the learners, their score, mistakes usually made by them etc. This framework facilitates the instructor in achieving this task as follows:

- Packaging teaching contents in a SCORM- compliant zip file called Package Interchange Format (PIF)
- The LMS enables content sharing and tracking of the learner by uploading this PIF file

IV. SCORM ENABLED LMS HYBRID FRAMEWORK

The macro level framework specified above can be elaborated further with component details. It has three layers namely database layer which is the base layer, business layer which is at the middle level, and client interface or the front end which forms the top layer. The limitations of existing LMS models identified through literature review and data analysis are addressed in this model. This framework has three parts, namely the database part containing the repository and teaching contents, LMS part which incorporates all the user requirements and modern features, and SCORM part which makes the LMS to be SCORM compliant or SCORM compatible

A. LMS Challenges and the Hybrid Framework

The new SCORM enabled hybrid framework shown in Figure - 4 is capable of addressing most of the challenges that the current LMS systems. It is evident from data analysis that an implementation gap exists in LMS systems. The major gaps were identified in the following aspects:

- Interoperability
- Accessibility
- Adaptability
- Reusability
- Durability
- Affordability

Integrating SCORM into the LMS system will provide a solution for many of the above mentioned challenges especially in terms of interoperability, accessibility and reusability.

1) Interoperability

The principal advantage of SCORM enabled hybrid LMS model is interoperability. Whenever an institution produces e-learning content, the clients will definitely demand to share it to an LMS. Similarly, when an organisation creates an LMS, invariably the clients will want to import learning content from a variety of sources into that LMS. This SCORM enabled hybrid LMS permits this incorporation to happen seamlessly. With a normal LMS, integrating with other vendors is a time-consuming and costly process. Thus the SCORM enabled hybrid LMS will make the institution more competent and lessen the support liability.

SCORM is considered as the de facto industry standard for interoperability in e-learning tools or LMS. SCORM conformance is a mandate in the e- learning space in order to make it interoperable. The learning contents which can play only in one place or platform has limited audience. Simultaneously an LMS that can play content particularly designed for that LMS also doesn't have much impact. Various tools integrated with this hybrid LMS enables the user to create Content that can be played 'anywhere'. This interoperability feature will definitely make this LMS more appealing to the customers.

2) Reusability

Reusability is the mantra of the software engineering industry today. Maximising the reusability feature will definitely minimise the cost and time. This principle is applicable in LMSs and e-content too. One of the greatest objectives and advantages of SCORM enabled hybrid LMS is that the instructional content, tools and applications once generated can be reused. This improves the quality, saves time and reduces the cost.

SCORM explicitly addresses the capacity to produce this large repository of learning contents. These learning contents should be small discrete chunks in order to make an automated system to be able to aggregate a customer tailored course. It should behave in a standard manner, and to be assembled in sequence. The training components are called SCOs. They have a common packaging scheme, common run-time behaviour; they are described by metadata and are put together in sequenced packages. Various tools and applications integrated with the system can also be reused along with the content.

3) Accessibility

Accessibility is the ability to locate and access instructional components from one remote location and deliver them to many other locations using a variety of devices working on various platforms. SCORM certainly does enable accessibility

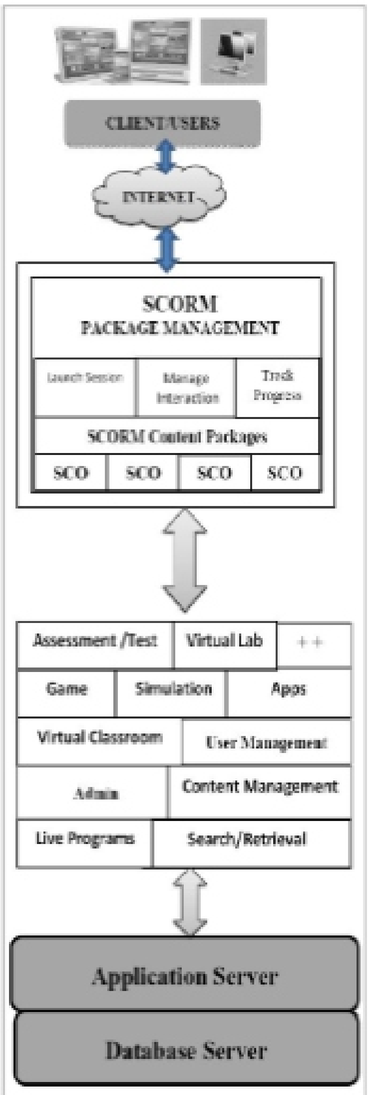


Figure 4: LMS SCORM Framework

This hybrid LMS is designed in such a way that it can meet almost all the accessibility standards. It is compatible with all the major browsers and accessible to the modern handheld devices such as tablets and smartphones. It functions as an application which facilitates online and offline mode of operations. SCORM courses are all packaged in a way that makes them easily portable across systems. A SCORM course can be delivered from any location without reconfiguration or complex installation. SCORM provides the technical framework for remote content to be catalogued and discovered.

4) Adaptability

Adaptability is the capability to shape instruction to individual and organizational needs. The modular structure of the hybrid LMS makes it possible to adapt and integrate any changes at any time without affecting the total structure. SCORM facilitates adaptability in two ways. It allows content authors to manually mix and match SCOs to produce distinctive training programs for diverse groups. It similarly permits them to compose sequencing rules that adapt the instructional content as the learner progresses based on the learner's input and mastery.

5) Affordability

Affordability is the capability to enhance the efficiency and productivity by reducing the time and costs involved in delivering instruction. SCORM has certainly reduced the costs of integrating content into LMS. The SCORM enabled hybrid LMS will certainly reduce the total cost because it can act as a single solution which addresses the entire ICT enabled requirements of an institution such as LMS, institution website, student management, staff management, communication tool, etc.

6) Durability

Durability is the ability to withstand technology evolution and changes without costly redesign, reconfiguration or recoding. The SCORM hybrid LMS can withstand the frequent changes in the software and hardware industry. SCORM has done a remarkable job of using established, mainstream technologies that will be widely supported and available for a long time.

B. SCORM Enabled LMS: Macro Level

A SCORM enabled LMS framework is built to address the challenges existing in the present ICT enabled teaching-learning process. This frame work addresses all the ICT requirements of an education institution under a single umbrella. The basic functionality of the package is to enhance the quality of teaching-learning process. It also caters other ICT requirements like online tests, staff evaluation, attendance, student management, staff management, parent-teacher interaction, institution website, various apps for teachers and students, integrations with social media, etc. under a single login. A macro level framework of the hybrid LMS implementation model is given in figure 5.

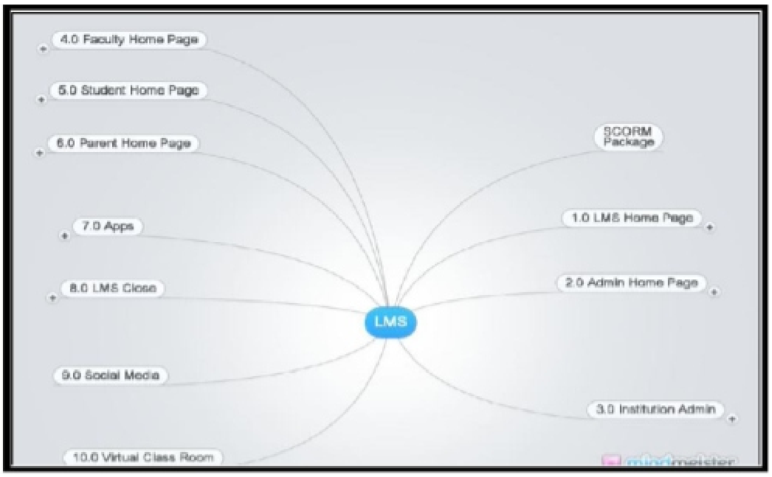


Figure .5: SCORM Enabled LMS Framework: Macro Level

C. SCORM Enabled LMS: Micro Level

Figure 5 gave the micro level implementation diagram of the proposed framework. The SCORM package is the central point of the LMS system.

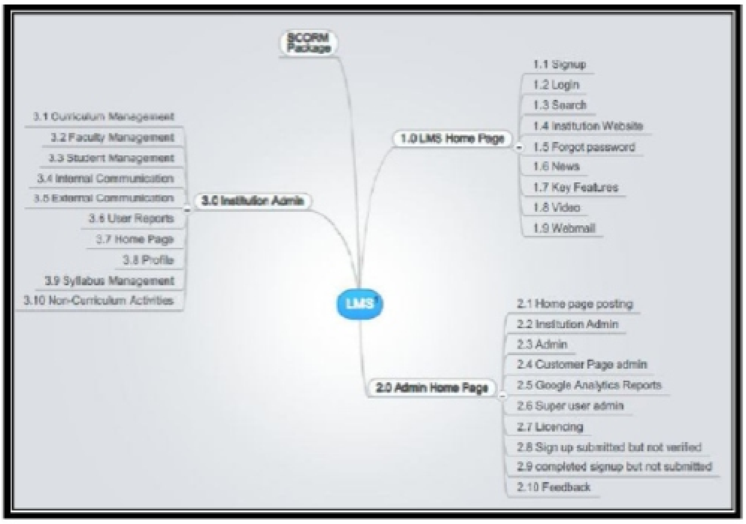


Figure 5: SCORM Enabled LMS Framework: Macro Level

CONCLUSION

The essential features of ICT based LMS systems were found and the implementation gap is identified. It is very evident that none of these parameters are completely addressed in the present LMS system. Enhancement of each parameter improves the total performance of the LMS system. A comparative study of various LMS systems pertaining to these parameters also shows the implementation gap of the essential LMS parameters such as interoperability, accessibility, adaptability, durability, reusability and affordability. The proposed hybrid LMS model would be an answer to these challenges and implementation of the same will definitely enhance the teaching- learning process.

REFERENCES

[1] O. and E. F. Ogbomo," ICT in Education: A Catalyst for Effective Use of Information", PNLA Quarterly Vol. 75, number 4, 2011.

[2] A. Riad and H.E.Ghareeb, "A service oriented architecture to integrate mobile assessment in learning management systems", Turkish online journal of distance education-TOJDE, April 2008.

[3] A. Riad and H.E.Ghareeb, "A service oriented architecture to integrate mobile assessment in learning management systems", Turkish online journal of distance education-TOJDE, April 2008.

[4] N.A Alias and A.M Zainuddin, "Innovation for Better Teaching and Learning: Adopting the Learning Management System." Malaysian Online Journal of Instructional Technology, Vol.2, 27-40, 2009.

[5] P. Baumgartner, H. Hafele, and K.H.Maier, E-Learning Praxishandbuch -Auswahl Von Lernplattformen. Studienverlag, Innsbruck, 2002

[6] A.Inaba and R.Mizoguchi, Learning Design Palette: An Ontology-aware Authoring System for Learning Design, ISIR, Osaka University, Japan 2008.

[7] O. Nwosu and E. F. Ogbomo," ICT in Education: A Catalyst for Effective Use of Information", PNLA Quarterly Vol. 75, number 4, 2011.

[8] P.Banking and S.Gallagher, "Choosing a Learning Management System", Advanced Distributed Learning (ADL) Co-Laboratories, Version 2.2.2, 2012.

[9] "AN INTRODUCTION TO THE SCORM BASICS", http://www.cybermediacreations.com/eleaning/white-papers.SCORM_101.pdf, April 2013.

[10] H.Srimathi and S.K. Srivatsa, "Design of Virtual Learning Environment Using SCORM Standards", Journal of Theoretical and Applied Information Technology, vol.2, 2008.

[11] P. Dodds, Advanced Distributed Learning, Sharable Content Object Reference Model, Version 1.2, The SCORM Content Aggregation Model, Available at ADLNET, http://www.adlnet.org, (Accessed 10th January 2013.)

[12] J.T. David Yang, C.H. Chiu & C.Yen Tsai, "Visualized Online Simple Sequencing Authoring Tool for SCORM-compliant Content Package", Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04), 2004.

[13] W. Bailey, "What is ADL SCORM", http://www.jatit.org/volumes/research-papers/vol4No6/12Vol4No6 .pdf, (Accessed on Feb. 2013).

[14] SCORM 2004 Hand Book, The e-Learning Consortium, Japan: 2006.

WEBRTC on Multi-Party Communication to Lower Video Streaming Traffic

[1] Yi-lun Chen, [2] Jeanne Chen, [3] Tung - Shou Chen
[1,2,3]Department of Computer Science and Information Engineering,
National Taichung Uiversity of Science and Technology
E-mail: 1s18023104@nutc.edu.tw, 2jeanne@nutc.edu.tw, 3tschen@nutc.edu.tw

ABSTRACT

Most mediastream servers use peer to peer (P2P) network infrastructure to resolve heavy load problem such as in WebRTC. WebRTC is characterized by a flexible signaling protocol method. To implement multi-party communication on this platform is relatively complex and difficult. In WebRTC, multi-party communication requires more bandwidth which increases by new peers. In this paper, we proposed a new WebRTC flow control mechanism, called the adaptive peer traffic control WebRTC (APTC WebRTC) for multi-party communications. Experimental results showed that APTC WebRTC can reduced the traffic for each peer in the multi-party communication to reduce local peer bandwidth hogging. At the same time, there is no video traffic consumption at the server-side of APTC WebRTC. This reduces the need for a large number of servers and costs are also recused.

Keywords— WebRTC, Peer to Peer, Multi-party Communication, Adaptive Traffic Control.

I. INTRODUCTION

Due to advancement in internet technology, most video communication software has been developed, such as Zoom, BigBlueButton (BBB), and WebRTC [1][3][4][5]. WEBRTC supports a variety of modules, which may be used to handle different network environment signaling connection management modules [8]. It does not provide signaling protocol specification. Therefore, it is combined with signaling protocol in order to improve compatibility. Furthermore, multi-party communication is difficult and complex to process. Examples of which include XMPP[6], SIP[2] and others.

There are several ways to use signaling protocol in WebRTC. Adeyeye et al. [9] implemented a three party communication for WebRTC. Signaling is accomplished by using SIP via WebSocket and JSON via XMLHttp Request (XHR). After comparing the overheads for these two methods, JSON via XHR showed less overhead. The current browser supports WebSocket using simple and easy method to achieve multi-party communication, such as JSON via WebSocket. Video and voice communications will generate increasing amounts of streaming data. In multi-party communications, these streaming data will consume a large amount of network bandwidth [7]. Ng. et. al. [10] proposed the P2P-Multipoint Control Unit (P2P-MCU) server method to share the bandwidth and CPU usage when each traditional WebRTC participates in the multi-party video chat. However, P2P-MCU results in some delays on the client side. The server requires higher computing resource and bigger network bandwidth.

When there is a new peer connection added to the network, the P2P-MCU server will require more CPU and network bandwidth.

This study proposed an adaptive peer flow control mechanism, called Adaptive Peer Traffic Control WebRTC (APTC WebRTC), to reduce the multi-party communication traffic consumption of traditional WebRTC, and to implement the signaling protocol through a javascript plugin. Comparisons on the multi-party communication traffic will be made with Zoom and BBB. This also includes the relationship between the number of peer connections and traffic consumption in APTC WebRTC.

II. THE PROPOSED APTC WEBRTC METHOD

WebRTC video chat is exchanging video streaming data between peers in the network. For the existing peers, newly added peers will consume more traffic for all peers.

In this paper, we propose an adaptive peer traffic control based multi-party communication WebRTC (APTC WebRTC). In APTC WebRTC, the network resolution and bandwidth are adjustable based on increase or decrease upload/download traffic. Fig. 1 shows the comparison diagram of bandwidth requirement for multi-party communication between the traditional WebRTC and APTC WebRTC. Each shows four peer connections. The arrow paths displayed the streaming data transmissions. The thicker arrow paths showed that the traditional WebRTC requires significantly larger bandwidth for its streaming data.

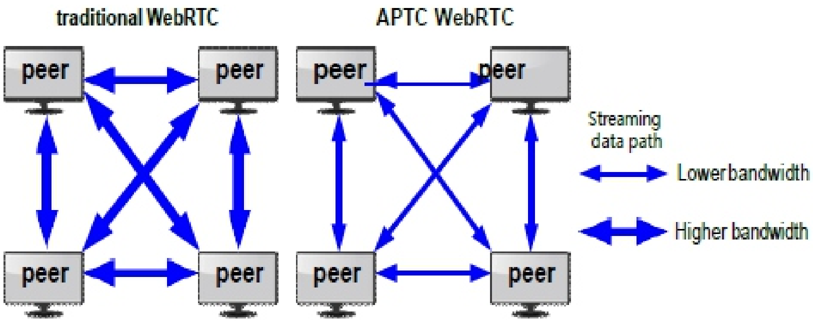


Fig. 1. Multi-party communication streaming data transmission.

The thinner arrow paths in APTC WebRTC imply smaller bandwidth requirement for the streaming data. The directional arrows pointing towards each peer will be noted as branches. For example, Fig. 1 showed the number of branches for each peer is 3 and the total upload/download bandwidth is the sum of the three upload/download bandwidth paths. As shown in Fig. 1, the number of branches in traditional WebRTC is increased to 3 by adding a new remote peer to the network. The total bandwidth of the local peer is increased three times from the origin. The changes in the bandwidth are as shown in Fig. 2.

Fig. 2. Traditional WebRTC local side bandwidth changes.

If remote peers are increased to n then the number of branches for each peers will be increased to n. When the value of n increases, the upload and download traffic will grow up to n times. The limited bandwidth network environment for general user, cannot afford high traffic consumption. Fig. 3 shows the diagram traditional WebRTC change to n remote peers.

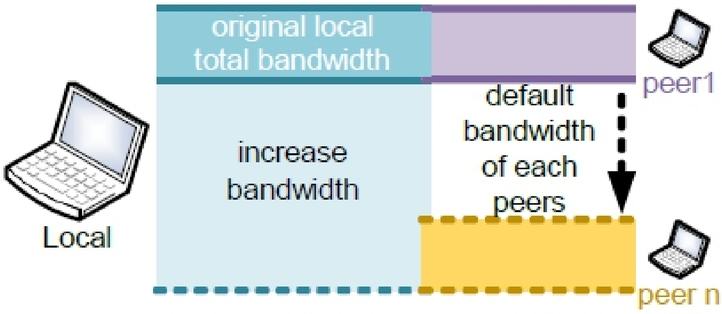


Fig. 3. Traditional WebRTC change to n remote peers.

Fig. 4 shows the diagram of APTC WebRTC changing bandwidth and resolution. Remote peer in (1) showed maximum resolution with one peer only. However, when (2) increases to two remote peers to start a multi-party communication, the resolution of each peer have to be readjusted.

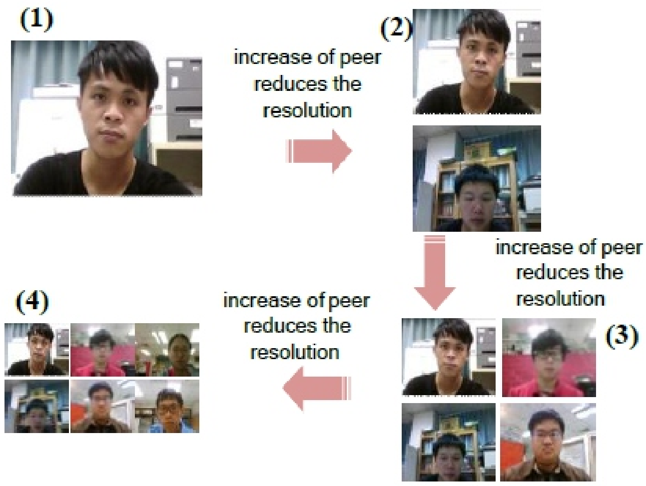


Fig. 4. Adjusting resolution of each remote peer with new peer addition.

After readjusting resolution, upload/download bandwidth is also reduced for each remote peer so as to reduce network traffic consumption. Fig. 5 shows APTC WebRTC adjustments of bandwidth when there are three remote peers.

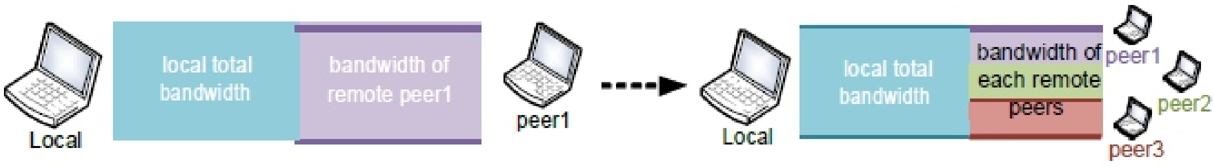


Fig. 5. APTC WebRTC adjustments of remote peer bandwidths with new peer additions.

Fig. 6 shows the system control flow for APTC WebRTC.

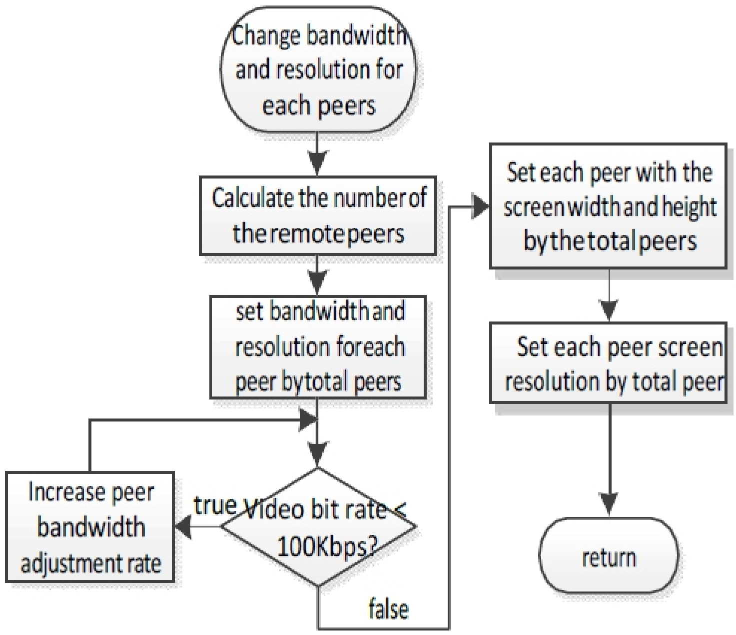


Fig. 6. Flow of adaptive peer traffic control WebRTC (APTC WebRTC).

When WebRTC is multi-party communication, the local peer traffic consumption is related to the branch of remote peers. Fig. 7 shows the relationship between traffic and branches. In comparison to the other methods, the line curve for traditional WebRTC (line tagged with triangles) reveals that it consumes the most amount of traffic. The diamond tagged line curve of trial APTC WebRTC is generated by the simulation with fixed traffic. The rectangle tagged line curve is the empirical APTC WebRTC results of possible traffic consumption in relation to the number of branches.

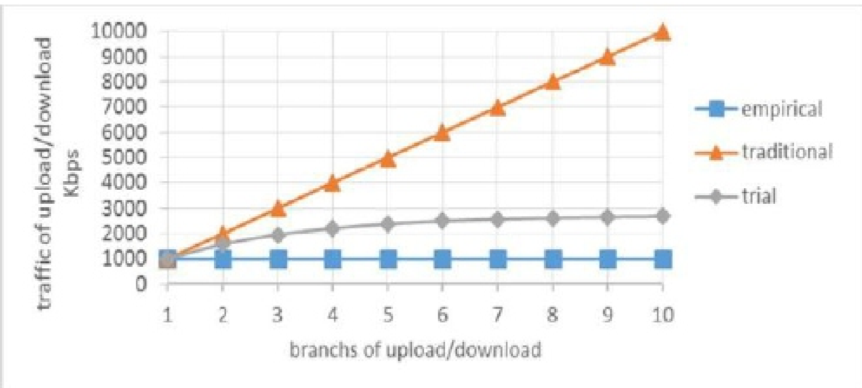


Fig. 7. Relationship between branches and bandwidths.

In this study, implementation of multi-party communication APTC WebRTC is based on RTC Multi-connection. Fig. 8 shows a flow chart for this system. The numbered markings show the sequence of processing the communication for each peer. (1) shows any peer request to Web server for web content. After the page is loaded to the peer, (2) shows the establishment of peer communication through the signaling server. When peer in (3) received the location of the remote peer, the streaming media transmission channel is established via a signaling protocol.

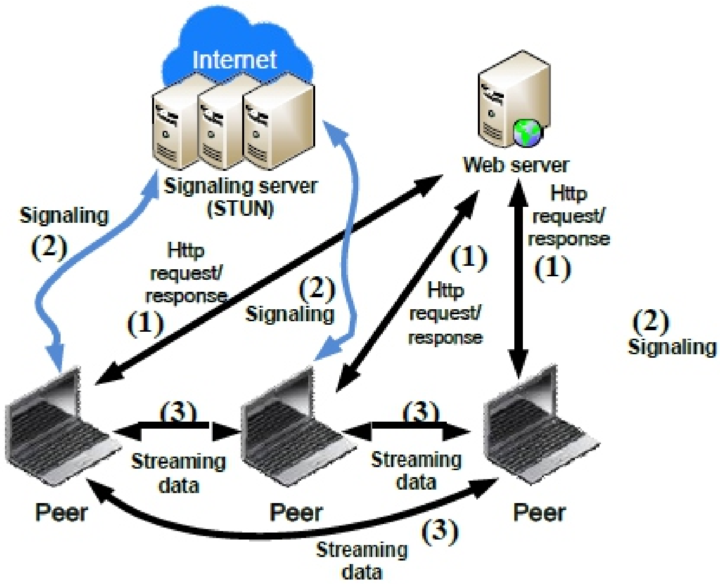


Fig. 8. System flow of APTC WebRTC.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this study, APTC WebRTC is the agent for video communication traffic consumption reduction via RTC Multiconnection library to handle signaling protocol of multi-party communication. Fig. 9 shows the communication process.

Fig. 9. APTC WebRTC multi-party communication.

In the experimental testing a total of five PCs were used and the network environment as shown in Fig. 10. The network environment is Ethernet network through switches connected to the router and then to the server.

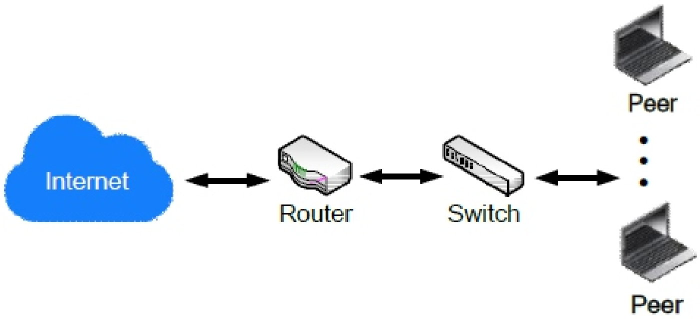


Fig. 10. Network experimental environment.

The experimental testing was divided into three parts. The test time is 30 minutes for each part. At different stages of the experiment each PC as a peer was added to the network to begin communication. First is the test for traffic consumption of peer participation in multi-party communication relationship between the number of communicating peers to traditional WebRTC and APTC WebRTC. Second is the test of the traffic consumption relationship to the number of peer participation in client-side of multi-party communication to Zoom, BBB and APTC WebRTC. Third and final part is the test traffic consumption relationship to the number of client connected at server-side. Part 2 and part 3 of experimental testing will be not performed for two peers. Therefore, testing starts at three peers.

Part 1 of the experimental results is as shown in Fig. 11. When communicating peers are increased to five, traditional WebRTC showed upload/download traffic increased to 5569 Kbps and 5558 Kbps. APTC WebRTC upload/download traffic was 1680Kbps and 1632Kbps because the number remote peers detected in the network requires adjustment to bandwidth and resolution. The new peer addition reduces traffic increment proportion. On the other hand the local peer has less total bandwidth change.

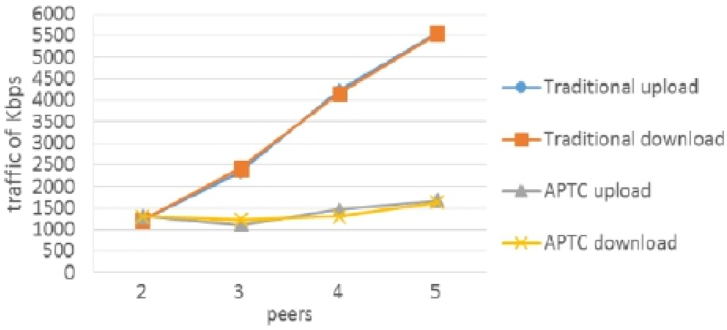


Fig. 11. Relationship between peers and traffic of multi-party communications for the traditional WebRTC vs APTC WebRTC.

Part 2 of peer and traffic consumption at client-side of Zoom, BBB and APTC WebRTC is as shown in Fig. 12. The number of the peer participation for Zoom, BBB and APTC WebRTC is increased from three to five with download traffic proportional changes as 49.36%, 125.28% and 33.44%. The total download bandwidth of APTC WebRTC is significantly lower than the other two.

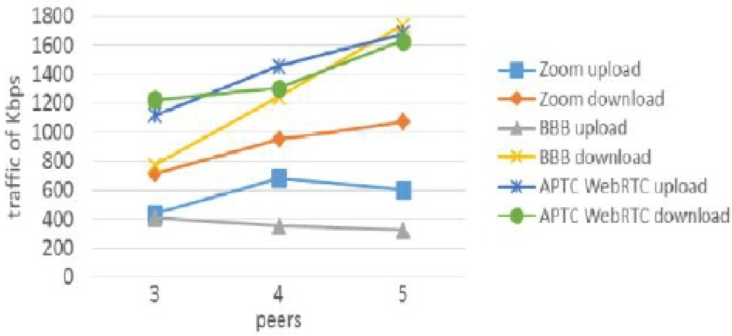


Fig. 12. Zoom, BBB and APTC WebRTC relation between multi-party communications and traffic.

The third and final part testing traffic relationship between the number of clients at server-side of Zoom, BBB and APTC WebRTC is as shown in Fig. 13. When clients of Zoom and BBB are increase to 5 the upload traffic was 5353Kbps, and 5213Kbps. This is due to the client-server based network which require transfer for the streaming data clients through the server. Peers content request/response or signaling packets pass through the server so that server side only has a small flow. The reason is due to APTC WebRTC based P2P network architecture.

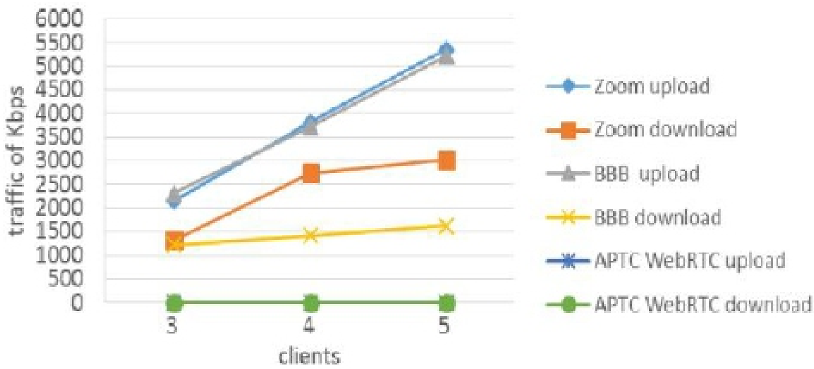


Fig. 13. The number of clients connected and traffic on the server-side.

APTC WebRTC improves traffic consumption of multi-party communication. There is less bandwidth increases when new peer is added to the network. APTC WebRTC has lower download increase proportion of traffic then Zoom and BBB. The server-side has less traffic consumption for the content request/response and signaling.

CONCLUSIONS

It is relatively difficult and complex to implement multi-party communication protocol for the traditional WebRTC. More connected peers required higher network bandwidth in the traditional WebRTC in multi-party communication.

This study proposed the APTC WebRTC to reduce the video chat traffic, and to process the multi-party communication signaling protocol via a javascript plugin. APTC WebRTC reduces network traffic by adjusting the resolution and bandwidth of all connected peers accordingly to the number of peers in the network. Experimental results showed that APTC WebRTC bandwidth consumption is significantly less than the traditional WebRTC at a local peer with additional peers. Comparisons with Zoom and BigBlueButton, when the client-side connected peers are increased from three to five, showed the proportion of traffic increase is much smaller than the other two for APTC WebRTC. APTC WebRTC generated less traffic consumption at server-side to reduce the device costs of server-side.

In conclusion, APTC WebRTC can reduce traffic consumption, and is suitable for use in multi-party video communication.

REFERENCES

[1] F. Rhinow, P. P. Veloso, C. Puyelo, S. Barrett and E. O. Nuallain, "P2P Live Video Streaming in WebRTC," in Computer Applications and Information Systems (WCCAIS), 2014 World Congress on, vol. 32, pp. 158-165, 2014, doi: 10.1109/WCCAIS.2014.6916588.

[2] J. Liao, J. Wang, T. Li, J. Wang, J. Wang and X. Zhu, "A distributed end-to-end overload control mechanism for networks of SIP servers," Computer Networks, vol. 56, Issue 12, pp. 2847-2868, 2012, doi:10.1016/j.comnet.2012.04.024.

[3] Google Inc., WebRTC, Google Inc., retrieved [online]: https://sites.google.com/site/webrtc/home, Sep, 2015.

[4] C. Togay, "WebRTC technology for mobile devices," in Signal Processing and Communications Applications Conference (SIU), 2014 22nd, vol. 51, Issue. 4, pp. 256-259, 2014, doi: 10.1109/SIU.2014.6830214.

[5] H. M. UK Ltd., Google open source WebRTC for open video/audio chat, retrieved [online]: http://www.h-online.com/open/news/item/Google-open-so urce-WebRTC-for-open-video-audio-chat-1253848.html, Jun, 2011.

[6] C. Cola and H. Valean, "On multi-user web conference using WebRTC," in System Theory, Control and Computing (ICSTCC), 2014 18th International Conference, pp. 430-433, 2014, doi: 10.1109/ICSTCC.20 14.6982454.

[7] W. Elleuch, "Models for multimedia conference between browsers based on WebRTC," in Wireless and Mobile Computing, Networking and Communications (WiMob), 2013 IEEE 9th International Conference on, pp. 279-284, 2013, doi: 10.1109/WiMOB.2013.6673373.

[8] J. H. Paik and D. H. Lee, "Scalable signaling protocol for Web real-time communication based on a distributed hash table," Computer Communications, vol. 70, pp. 28–39, 2015, doi:10.1016/j.comcom.2015 .05.013.

[9] M. Adeyeye, I. Makitla and T. Fogwill, "Determining the signalling overhead of two common WebRTC methods: JSON via XMLHttpRequest and SIP over WebSocket," in AFRICON, 2013, pp. 1-5, 2013, doi: 10.1109/AFRCON.2013.6757840.

[10] K. F. Ng, M. Y. Ching, Y. Liu, T. Cai, L. Li and W. Chou, "A P2P-MCU Approach to Multi-Party Video Conference with WebRTC," IJFCC 2014, vol. 3, no. 5, pp. 319-324, 2014, doi: 10.7763/IJFCC.2014.V3.319.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

-

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

<u>**Address of the Editorial Office:**</u>

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Note