المؤتمر الليبي الدولي لتقنيات المعلومات والاتصالات

تحت شعار

سبل التحول الى الاقتصاد المعرفي

The International Libyan Conference for Information and Communications Technologies (ILCICT 2022)

Under the Slogan

Ways to Transition to A Knowledge Economy

تحت اشراف وتنظيم

كلية الهندسة - جامعة طرابلس

Organized and Supervised by

Faculty of Engineering – University of Tripoli

أحد فاعليات الدورة الثانية

لملتقى ليبيا الدولي لتكنولوجيا المعلومات

One of the activities of the second session of

The Libya International Forum for Information Technology

كتاب أبحاث المؤتمر

CONFERENCE PROCEEDINGS

طرابلس – ليبيا 27 – 30 مارس 2022

Tripoli – Libya 27 – 30 March 2022

https://lit.ly/ILCICT2021/

Table of Contents جدول المحتويات

Preparatory Committee — اللجنة التحضيرية

تم الإعداد والتحضير لهذا المؤتمر من قبل اللجنة التحضيرية الدورة الثانية لملتقى ليبيا الدولي لتكنولوجيا المعلومات

This conference has been organized by the organization committee of the second session of The Libya International Forum for Information Technology

Scientific Committee — اللجنة العلمية للمؤتمر

Chair — رئيس اللجنة

Dr. Mohamed Samir Elbuni — **د. محمد سمير عبد الرحمن البوني**
Computer Engineering Department — قسم هندسة الحاسب
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

Committee Members — أعضاء اللجنة

Dr. Ibrahim M. Saleh — **د. إبراهيم محمد صالح**
Electrical and Electronic Engineering Department — قسم الهندسة الكهربائية والإلكترونية
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

Dr. Ahmed S. Ashur — **د. أحمد سعيد عاشور**
Electrical and Electronic Engineering Department — قسم الهندسة الكهربائية والإلكترونية
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

Dr. Tammam A. T. Benmusa — **د. تمام أحمد تيسير بن موسى**
Electrical and Electronic Engineering Department — قسم الهندسة الكهربائية والإلكترونية
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

Dr. Issmail M. Ellabib — **د. أسماعيل محمد لبيب**
Computer Engineering Department — قسم هندسة الحاسب
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

Dr. Nabil M. Drawil — **د. نبيل محمد دراويل**
Computer Engineering Department — قسم هندسة الحاسب
Faculty of Engineering – University of Tripoli — كلية الهندسة - جامعة طرابلس

International Reviewer Committee لجنة المحكمين الدولية

Abdulatif Khrwat	Faculty of Engineering, Al-Jabal AlGharbi University, LIbya
Abdulbaset Hamed	Western University, Canada
Abduljalil Mohamed	Ahmed Bin Mohammed Military Collage
Abdulmagid Aborwin	Irvine Valley College, USA
Abdurhman Albasir	Praemo Inc., Canada
Abubaker Abushofa	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Adel Saad	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Ahmad Kharaz	The School of Engineering and Technology, University of Derby, UK
Ahmed Alkilany	Faculty of ICT, University of Tripoli, Libya
Ahmed Abougarair	University of Tripoli, Libya
Ahmed Kagilik	Sabratha University, Libya
Akrem El-ghazal	Conestoga College Institute of Technology, Kitchener, Ontario
Ali Ganoun	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
ALI ELGHARIANI	XCOM Labs
Almahdi Ibrahim	Fazzan University, Libya
Amer ZEREK	Zawia University, Libya
Amna Elhawil	Computer Eng. Dept. University of Tripoli, Libya
Ayad Ali	RIC
Azeddien Sllame	Faculty of ICT, University of Tripoli, Libya
Esmail Abuhdima	Embry-Riddle Aeronautical University, USA
Haitham Amar	University of Waterloo, Canda
Husam Alfergani	Rowan University, USA
Ibrahim Lahmer	National Oil Corporation
Ismail Shrena	Faculty of ICT, University of Tripoli, Libya
Khaled Shaban	Qatar University, Qatar
Khaled Dadesh	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Labib Daloub	ACECS
Lutfi Albasha	American University of Sharjah, UAE
Mahdi Hafi	Computer Eng. Dept. University of Tripoli, Libya
Majdi Ashibani	Libya Telecom and informatics Academy, Libya
Mohamed Twati	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Mohamed Elalem	Elmergib University, Libya
Mohamed Elmansouri	University of Colorado Boulder, USA
Mohamed Eltarhuni	American University of Sharjah, UAE
Mohamed Mussa	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Mohamed Eljhani	Computer Eng. Dept. University of Tripoli, Libya
Mohamed Elgalhud	Computer Eng. Dept. University of Tripoli, Libya
Mohamed Ghretli	College of Computer Technology Tripoli, Libya
Mohamed Sbeta	Libyan Center for Solar Energy Research and studies, Libya
Mrwan Margem	College of Electronic Technology Tripoli, Libya

Mustafa Alasswad	Faculty of Engineering, Sebrata University
Nabil Naas	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Nizar Khemri	Computer Eng. Dept. University of Tripoli, Libya
Nuri Benbarka	University of Tübingen, Germany
Omar Bouzid	Faculty of Engineering, Al-Jabal AlGharbi University, LIbya
Osama Alkishriwo	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Said Elkhetali	College of Electronic Technology, Bani Walid
Salah Kanoun	Computer Eng. Dept. University of Tripoli, Libya
Sami Bizzan	Electrical and Electronic Eng. Dept. University of Tripoli, Libya
Sarra Elrabiei	Computer Eng. Dept. University of Tripoli, Libya
Seddeq Ghrare	Faculty of Engineering, Al-Jabal AlGharbi University, LIbya
Suad El-Geder	Computer Eng. Dept. University of Tripoli, Libya
Tarek El-mihoub	Jade university for applied science
Tarek Khalifa	American University of the Middle East, Kuwait
Yusra Maatug	Computer Eng. Dept. University of Tripoli, Libya

تقديم اللجنة العلمية للمؤتمر

ينعقد المؤتمر الليبي الدولي لتقنية المعلومات والاتصالات في دورته الأولى تحت شعار سبل التحول الى الاقتصاد المعرفي وذلك خلال الفترة من 27 إلى 30 مارس 2022 تحت إشراف وتنظيم كلية الهندسة جامعة طرابلس ورعاية ودعم القابضة للاتصالات والشركات التابعة لها, والهيئة العامة للاتصالات والمعلوماتية، وكذلك بمساهمة وتعاون بعض الجهات الأخرى ذات العلاقة.

وتتلخص أهم أهداف هذا المؤتمر فيما يلي :-

1) توفير ملتقى يجتمع فيه المهندسون والباحثون وخبراء الصناعة والاقتصاد المهتمون بمجال الاتصالات وتقنية المعلومات لتبادل الأفكار العلمية والبحثية ومتابعة كل جديد في التخصص في منتدى مفتوح.
2) تشجيع الباحثين في مجال الاتصالات وتقنية المعلومات على اجراء أبحاث علمية تخدم المجتمع وتساهم في تنمية الاقتصاد الوطني.
3) تشجيع الباحثين في مجال الاتصالات وتقنية المعلومات على نشر ابحاثهم العلمية.

وحرصاً على ظهور المؤتمر بالمظهر العلمي والحضاري اللائق في ليبيا وأن يحظى بسمعة طيبة في الأوساط العلمية الدولية فقد تم تشكيل لجنة محكمين علمية دولية تضم عدد من العلماء والأخصائيين الليبيين وغيرهم من العاملين بالجامعات والمعاهد والهيئات العلمية في ليبيا وعدداً من العلماء والأخصائيين المعروفين من جامعات وهيئات علمية عالمية وعربية. وقامت هذه اللجنة بإعداد البرنامج العلمي للمؤتمر وتقييم وتحكيم الأبحاث المقدمة واختيار أفضلها لإلقائها في المؤتمر والنشر في كتاب الأبحاث الصادر عنه. كما قامت بالتجهيز لمحاضرات علمية يتم فيها التعرض لأحدث التقنيات وآخر ما توصل اليه العلم في مواضيع ذات علاقة بأهم مجالات المؤتمر بالإضافة الى التنسيق لعدد من حلقات النقاش تتعلق بأهم قضايا الاتصالات وتقنية المعلومات في الدولة الليبية.

اعتمدت اللجنة العلمية برنامج Conference Management System CMT وهو أحد برمجيات شركة ميكروسوفت والذي يعتبر أحد أشهر البرمجيات الخاصة بتنظيم وإدارة المؤتمرات العلمية ويساعد في التنسيق في تقييم الأبحاث دون معرفة المؤلفين، وعملت اللجنة العلمية على تقييم الأبحاث بتحويل كل بحث إلى اثنين من المراجعين المختصين من قائمة المقيمين من أجل تقييمها وتحكيمها وفى حالة اختلافهما يتم تحويلها إلى مقيم ثالث، للبث في وضع الورقة من حيث القبول أو الرفض.

وقد وضعت اللجنة العلمية عدة ضوابط وشروط لتقديم الأوراق البحثية مماثلة للضوابط المعمول بها في المؤتمرات العالمية المشابهة، وتم فتح باب قبول الورقات البحثية في عدد خمس محاور رئيسية تشمل عدة نقاط لكل محور على النحو التالي:-

المحور الاول: انظمة الحاسوب والمعلوماتية
المحور الثاني: أنظمة الاتصالات
المحور الثالث: أنظمة أمن وسلامة المعلومات
المحور الرابع: الشبكات
المحور الخامس: تطبيقات في تطوير وتنمية الاقتصاد الوطني

وقد تم استلام 51 ورقة بحثية من أساتذة وخبراء ومختصين من الجامعات والمعاهد ومراكز الأبحاث المحلين والعالميين وتم استبعاد 6 اوراق نتيجة لعدم استفاء الشروط المطلوبة وتمت إحالة 45 ورقة للجنة العلمية العالمية للمحكمين ونظراً لتغيير موعد انعقاد المؤتمر من شهر 9 / 2021 الى شهر 3 / 2022 فقد تم سحب 5 ورقات أخرى ، وبناءً على رأي السادة المحكمين فقد تمت إجازة 24 ورقة فقط للإلقاء في المؤتمر والنشر في كتاب الأبحاث هذا وتم رفض 16 ورقة ، أي أن نسبة الاجازة كانت 60 % من الأوراق المحالة للتحكيم و47 % من اجمالي الأوراق المستلمة ، وقد كان توزيع هذه الأوراق على المحاور على النحو التالي :-

المحور	الاتصالات	الحاسب	الشبكات	أمن المعلومات	دعم الاقتصاد الوطني	الاجمالي
اجمالي الورقات المستلمة	10	26	6	3	6	**51**
تم استبعادها لعدم استيفائها شروط التقديم	2	2	0	0	2	**6**
المكتملة والتي أحيلت للتحكيم	8	24	6	3	4	**45**
ورقات تم سحبها من قبل المؤلفين	0	4	1	0	0	**5**
الورقات المقبولة	6	12	4	1	1	**24**
الورقات المرفوضة	2	8	1	2	3	**16**
الورقات المقبولة قبول مشروط	0	0	0	0	0	**0**

وتنشر البحوث في هذا الكتاب تماما كما أرسلها المؤلفون بعد إدخال ملاحظات المراجعين، وعليه فالمؤلف يتحمل مسئولية أية أخطاء مطبعية أو إملائية أو لغوية أو تنظيمية قد تظهر في بحثه. ويسر اللجنة العلمية أن تتقبل أية اقتراحات أو ملاحظات من شأنها أن تساعد في تحقيق الأهداف المرجوة للمؤتمر والوصول بالمؤتمرات القادمة إلى الأفضل.

وفى الختام يسر اللجنة التحضيرية لملتقى ليبيا الدولي لتكنولوجيا المعلومات واللجنة العلمية للمؤتمر باسم كلية الهندسة، وكلية الهندسة جامعة طرابلس والجهات الراعية للمؤتمر أن تشكر جميع أعضاء قائمة المراجعين لمجهودهم الذي بذلوه في مراجعة جميع البحوث التي قُدمت للمؤتمر.

كما يسر اللجنة العلمية أن تقدم خالص شكرها وتقديرها إلى كل من اللجنة التحضيرية لملتقى ليبيا الدولي لتكنولوجيا المعلومات والهيئة العامة للاتصالات والمعلوماتية الشركة الليبية القابضة للبريد والاتصالات وتقنية المعلومات والشركات التابعة لها, على رعايتهم ودعمهم المستمر للمؤتمر وجميع الجهات الأخرى التي ساهمت في إنجاح المؤتمر. كما يسر اللجنة أن تشكر جميع أعضاء اللجان العاملة للمؤتمر وجميع الذين شاركوا وساهموا في جميع نشاطات المؤتمر.

هذا المؤتمر هو المؤتمر الأول من سلسة مؤتمرات ستعقد بإذن الله سنويا، أي كل سنة وتدعوكم اللجنة التحضيرية واللجنة العلمية لمواصلة المشاركة في اللقاءات القادمة.

اللجنة العلمية

Submission of the Scientific Committee to the Conference

The International Libyan Conference for Information and Communications Technologies (**ILCICT**) will be held in its first session on the theme "Ways to transform to a knowledge economy" from 27 to 30 March 2022 under the supervision and organization of Faculty of Engineering at University of Tripoli and the sponsorship and support of the holding of communications and its subsidiaries.

The main objectives of this Conference are:

1. Provide a forum for engineers, researchers, industry and economists interested in communications and information technology to exchange scientific and research ideas and to pursue new developments in an open forum.
2. Encouraging researchers in communications and information technology to carry out scientific research that serves society and contributes to the development of the national economy.
3. Encourage researchers in communications and information technology to disseminate their scientific research.

In order to ensure that the Conference has a sound scientific and civilized appearance in Libya and to attain a good reputation in the international scientific community, an International Scientific Review Committee has been set up, comprising a number of Libyan scientists, engineers and specialists working at universities, institutes and scientific bodies in Libya, in addition to other well-known figures from international universities and Arab scientific bodies. This committee prepared the scientific Conference program, evaluated, reviewed and selected the best research papers to be presented at the Conference and published in its proceedings. Furthermore, it has prepared scientific lectures to present the latest technologies and scientific findings on topics related to the most important areas of the Conference. The committee has also coordinated a number of panel discussions on the most important issues related to communication and information technology in the State of Libya.

The Scientific Committee adopted Microsoft Conference Management System CMT which is one of the most popular software in the area of organization and management of scientific conferences. This software helps coordinate the evaluation of research papers without the knowledge of authors. The Scientific Committee has evaluated the research papers through forwarding each paper to two relevant reviewers from the evaluators' roster for evaluation and revision. If their evaluations were different the paper is submitted to a third reviewer for final evaluation related to acceptance or rejection.

The Scientific Committee has established several conditions and requirements for submission of research papers according to the best international practices. The research papers have been opened for acceptance in five main areas, each covering several research points as follows:

Computers and Informatics systems

Communication Systems

Information security

Network

Applications in developing the national economy

Fifty-one papers were received from professors, experts and specialists working in local and international universities, institutes and research centers. Six papers were excluded as a result of non-compliance. 45 papers were forwarded to the International Scientific Review Committee. In view of the change in the date of the Conference from 9/2021 to 3/2022, five more papers were withdrawn. According to the opinion of the reviewers, only 24 papers were cleared for the Conference and published in its proceedings and 16 papers were rejected, i.e. 60% of the papers submitted for revision and 47% of the total papers received.

Theme	Communications	Computer	Networks	Information Security	National Economy Support	Total
Total papers received	10	26	6	3	6	**51**
Disqualified for not complying with application requirements	2	2	0	0	2	**6**
Completed and transferred to arbitration	8	24	6	3	4	**45**
Papers withdrawn by authors	0	4	1	0	0	**5**
Accepted Papers	6	12	4	1	1	**24**
Rejected Papers	2	8	1	2	3	**16**

The research papers in the proceedings is published exactly as sent by the authors after the introduction of the auditors' observations, and the author is therefore responsible for any typographical, spelling, linguistic or organizational errors that may appear in his manuscript. The Scientific Committee is pleased to receive any suggestions or observations that could help to achieve the Conference's objectives and to contribute in the success of future conferences.

Finally, the Organizing Committee for the Libya International Meeting on Information Technology, the Scientific Committee of the Conference on behalf of the Faculty of Engineering, the Faculty of Engineering of the University of Tripoli and the sponsors of the Conference are pleased to thank all members of the List of Auditors for their efforts in reviewing all the research submitted to the Conference.

The Scientific Committee is also pleased to extend its sincere thanks and appreciation to both the Organizing Committee for the Libya International Information Technology Forum and the General Authority for Communications and Information Technology, the Libyan Post Telecommunication and Information Technology (Holding Company) and its subsidiaries, for their continued sponsorship and support of the Conference and all other entities that contributed to the Conference's success. The Scientific Committee is also pleased to thank all members of the

Conference's working committees and all those who have participated in and contributed to all the activities of the Conference.

This Conference is the first of a series of conferences to be held yearly and the Organizing Committee and the Scientific Committee invite you to continue to participate in the forthcoming meetings.

The Scientific Committee

الأوراق البحثية

PAPERS

Voltage Reference Circuits Comparisons in 65nm CMOS Process

Alharari alsouri Alharari
EEE Department
University of Tripoli
Tripoli, Libya
al.alharari@uot.edu.ly
alhararyalsoory@gmail.com

Sami Saddek Bizzan
EEE Department
University of Tripoli
Tripoli, Libya
s.bizzan@uot.edu.ly
sbizzan@gmail.com

Abdulmoied Omar
EEE Department
University of Tripoli
Tripoli, Libya
aa.omar@uot.edu.ly
abdalmoed92@gmail.com

Abstract— Many voltage reference circuits found in the literature claim superior performance over their peers despite the fact that each result is obtained with the use of different process parameters. This paper attempts to provide some means of comparing these circuits in the same process node. Selected CMOS voltage reference circuits are re-designed (re-sized only) and simulated with the same process parameters namely, 65nm PTM process. Design I gave the best overall results (temperature variation = 3.42ppm/cº and PSRR = -131dB) while design V operated with the lowest supply voltage (0.3V).

Keywords— Bandgap Circuits, Subthreshold Voltage Reference Circuits, Temperature Compensated Reference Circuits, Bias Circuits.

I. INTRODUCTION

Bandgap voltage reference is one of the most important circuit blocks in high performance systems. It is widely used in most modern integrated circuits (ICs). For example, it provides a reference voltage in digital to analog converters (DAC) or in analog to digital converters (ADC). The reference voltage also is used in various analog blocks such as phase locked loops (PLL) and crystal oscillators to generate bias currents.

Recent advances in integrated circuit processing and manufacturing such as increasing the number and types of devices available for use allowed for more sophisticated circuits to generate voltage references. Nowadays, a wide range of control systems, data processors, and power electronics use voltage reference circuits as an integral part of their design, which means that the voltage reference circuits appear in almost all electronic systems.

David Hilbiber presented one of the first bandgap voltage reference circuit in 1964 [2]. The proposed circuit outputs a voltage of 1.25V and achieves a temperature coefficient of 10 ppm/°C, 50 ppm/°C. Robert Widlar proposed a reference circuit in 1971 which later became known as the Widlar bandgap voltage reference circuit [3]. The circuit generates a stable low temperature coefficient reference voltage at (1.23V) and implemented in bipolar process technology [4]. Kuijk proposed a reference circuit 1973, the Kuijk bandgap voltage reference circuit implemented with the conventional bipolar technology [4].

Paul Brokaw introduced the well-known bandgap voltage reference (BGR), which resolve the issue of variability in the base current due to processing and temperature effects on beta in 1974 [5]. The previous circuit techniques will no longer be suitable, because of the high output voltage (about 1.25V). This problem can be solved by resistive voltage dividers to obtain a small output voltage. Neuteboom et al. proposed a reference circuit, which is based on the resistive division technique and provides an output voltage lower than 1.23V while the supply voltage is limited to 0.9V in 1997 [6]. Banba proposed a technique to overcome the limitation in lowering the supply voltage. Sub-1V bandgap voltage reference circuits in CMOS technology can be designed by using resistive sub-division technique to reduce the minimum required supply voltage [7].

It is clear that recent trends in designing voltage reference circuits are moving away from the traditional designs which implement bipolar transistors and instead, use CMOS transistors operate in subthreshold voltage region. This trend is motivated by two factors: wide spread use of CMOS process and decreasing supply voltage.

The objective of this paper is to present fair performance comparisons for various voltage reference circuits commonly used in the literature. The rest of the paper includes the following: section II discusses the operating principle of bandgap circuits, followed by a review of various voltage reference circuit designs in section III. Simulation and results comparisons for the reviewed circuits are provided in section IV and V respectively. Finally, a conclusion is drawn in section VII.

II. OPERATING PRINCIPLE OF BANDGAP CIRCUIT

The classical bandgap circuit operating principle relies on generating a voltage that increases linearly with temperature and it is called Proportional-To-Absolute-Temperature, PTAT. It also generates a voltage that decreases linearly with temperature and it is called Complementary-To-Absolute Temperature, CTAT. These two voltages are added together to obtain a voltage which is constant with respect to temperature variations across the desired range.

For MOS transistor operating in subthreshold voltage region where $V_{GS} \leq V_{TH}$, the gate to source voltage, V_{GS}, has a linear relationship with temperature, as shown by Equation (1),

$$V_{GS} = V_{GS}(T_0) - K_G\left(\frac{T}{T_0} - 1\right) \quad (1)$$

Where $K_G = K_T + V_{GS}(T_0) - V_{TH}(T_0) - V_{OFF}$ is constant [9]. The above equation along with some circuit technique is used to produce CTAT component.

Again, MOS transistor operating in subthreshold voltage region where $V_{GS} \leq V_{TH}$, the drain current exhibits exponential relationship with gate to source voltage as given by Equation 2,

$$I_D = I_0 \frac{W}{L} exp\left(\frac{V_{GS}}{\xi V_T}\right) \quad (1)$$

Where ξ is non-ideal factor and greater than 1 and $V_T = KT/q$. Implementing circuit technique to push different current densities into two MOS transistors yields a voltage difference between the gates that is PTAT component.

III. Various Voltage Reference Circuits

In this section, we will discuss different bandgap voltage reference circuits commonly found in the literature based on their working principle, implementation, and performance parameters. These circuits will be resized and simulated with the same CMOS process parameters (models) for comparisons.

A. ZTC Point Voltage Reference (Design I)

A sub-1V voltage reference exploiting the zero temperature coefficient (ZTC) point of a MOSFET operated in the suprathreshold region and hence, its characteristics is largely unaffected by junction leakage current. It is implemented in 65nm CMOS technology and the reference circuit is illustrated in Figure 1. This design also fine tune the ZTC point of the MOSFET by adjusting drain to source voltage to reduce gate to source voltage variation across the desired temperature range [8].

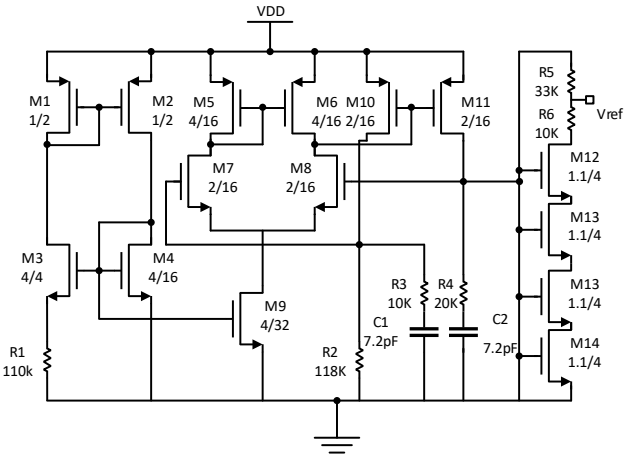


Figure 1: ZTC CMOS Voltage Reference (Desgn I)

B. Voltage Reference with Channel Length Modulation Compensation (Design II)

This is a simple voltage reference circuit with channel length modulation compensation. It is implemented in TSMC 0.18µm process technology, and the circuit is shown in Figure 2. The working principle of this circuit follows the traditional technique of pushing equal currents, using current mirrors, into different size nMOS transistors operate in the subthreshold region. The circuit can be divided into three parts. The first part includes transistor M1 to M5 along with resistor R1 and it generates a current proportional-to-absolute temperature (PTAT). The second part made of transistors M6 to M9 and resistor R2, which generates current complementary-to- absolute temperature (CTAT). The last part made of transistors M10, M11 and resistor R3, which responsible for adding both currents with proper ratio and converting it to voltage through the resistor R3 to obtain the reference voltage [9].

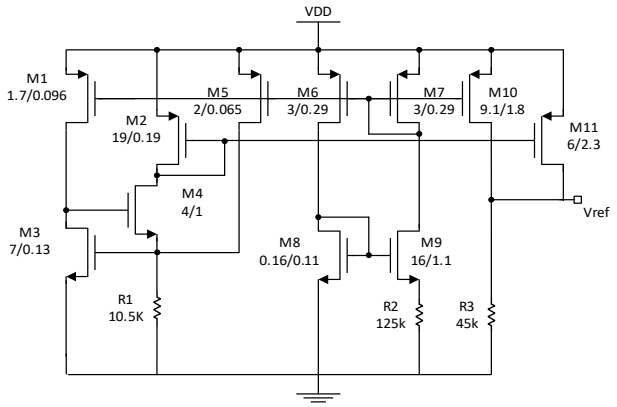


Figure 2: Voltage Reference with Channel Length Modulation Compensation (Design II)

C. Low Power Voltage Reference (Design III)

This design is implemented in 45nm CMOS process and the reference circuit is shown in Figure 3. Transistors M3 through M6 along with resistor R2 provide current proportional-to-absolute temperature (PTAT). Transistors M7 and M8 along with resistor R3 provide current complementary-to-absolute temperature (CTAT). The reference voltage is obtained by summing the two currents (through M1 and M2) at the output node and converting it to voltage with resistor R1 [10]. This design is very similar to design II above and is included since the circuit topology is very practical in terms of size and performance. This design also provides results for different processing technology node.

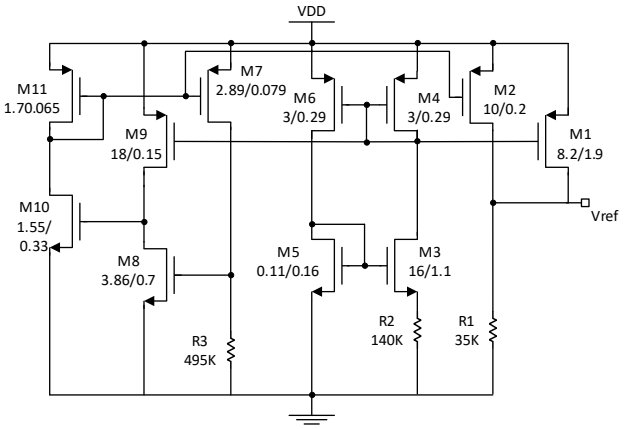


Figure 3: Low Power Voltage Reference (Design III)

D. Temperature and Supply Compensated Voltage Reference (Design IV)

This design utilizes a very simple OPAMP and startup circuit, and implemented in 45nm process technology. The reference voltage circuit is shown in Figure 4. It can be divided five functional parts. Part one includes transistors M4, M5, M10, M11, and resistor R2 which generate current proportional-to-absolute temperature (PTAT). Part two includes transistors M15, M16, M17, and resistor R3 which generate current complementary-to-absolute temperature (CTAT). Part three is the operational amplifier made of transistors M6, M7, M8, and M9 which provides a feedback structure. Part four is basically a startup circuit made of transistors M12, M13, and M14. Part five adds the two

currents at the output node and converts the total current into voltage through resistor R1 [11].

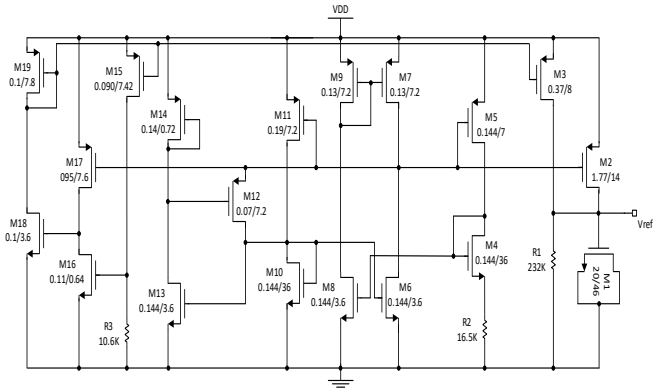


Figure 4: Temperature and Supply Compensated Voltage Reference (Design IV)

E. Self-Bised Voltage Reference (Design V)

A self-biased subthreshold voltage reference circuit implemented on a 0.18 µm CMOS process technology is shown in Figure 5. This design is realized by using a thin oxide MOS in series with a thick oxide MOS (M2) to produce reference voltage. The thick oxide MOS drain is selected as the output with optimized bias current [12]. For the purpose of comparison, this design is re-simulated with standard thin oxide transistors and with the sizes shown on the figure.

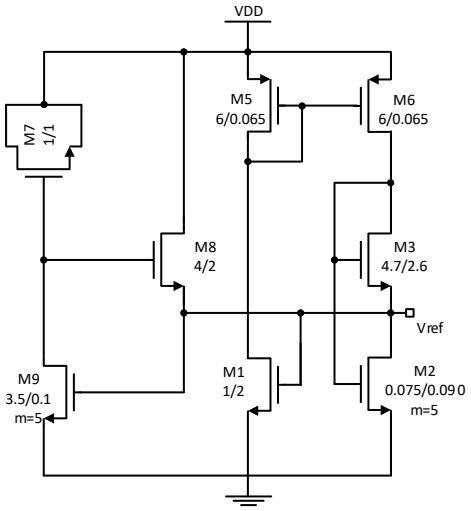


Figure 5: Self-biased Voltage Reference (Design V)

IV. SIMULATION AND RESULTS

Designs I through V have been re-designed (re-sized only) and simulated with the following setup,

- LT SPICE is used for the simulation.
- 65nm PTM publicly available models are used [1].
- Original transistor sizes are scaled up or down to suit the process at hand while keeping the relative sizing as close as possible to the original design.
- Supply voltage is set to 1.1V as required by the process.

Figure 6 through Figure 20 shows the simulation results for the circuits under investigation. For each design, there are three types of curves generated from the simulation, namely, reference voltage vs temperature, reference voltage vs supply voltage, and power supply rejection ratio vs frequency. The family of curves are generated by varying the supply voltage within the valid operation range as given by Table1 .

Design I shows the best results in all aspects and it is clearly well optimized with temperature coefficient of 3.42 ppm/°c. However, it occupies the largest silicon area as predicted by the gate area figure. It is worth mentioning that design V works fine and the reference voltage is stable for extremely low supply voltage (0.3V). This result is similar to the original design result reported in the literature.

Table1 is divided into two parts: the top part shows the results as reported in the literature for each design while the bottom part shows the results for each design simulated in 65nm process. For the purpose of generating power supply rejection ratio values at selected frequencies, we have set the supply voltage to minimum plus 10%. This will yield more practical figures as one never deploys the designs with minimum supply voltage.

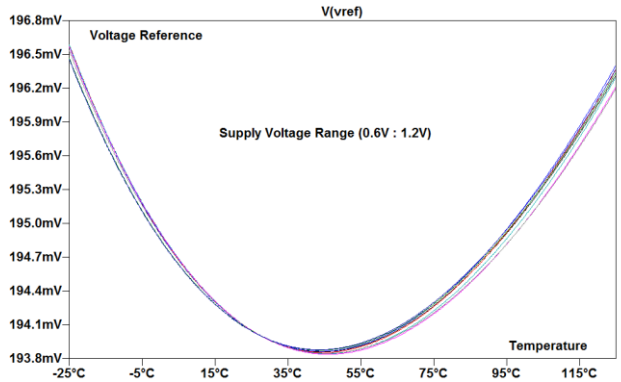


Figure 6: Voltage Reference vs Temperature (Design I)

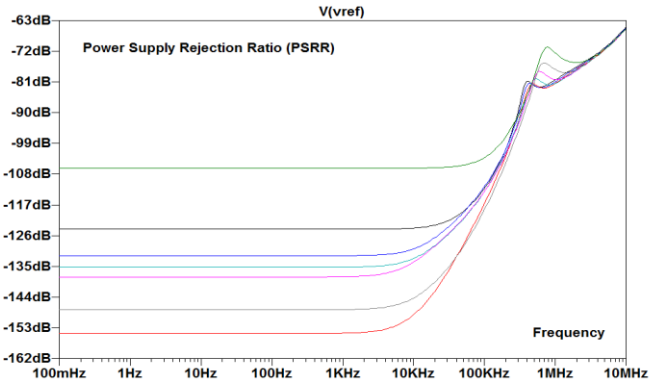


Figure 7: Power Supply Rejection Ratio vs Frequency (Design I)

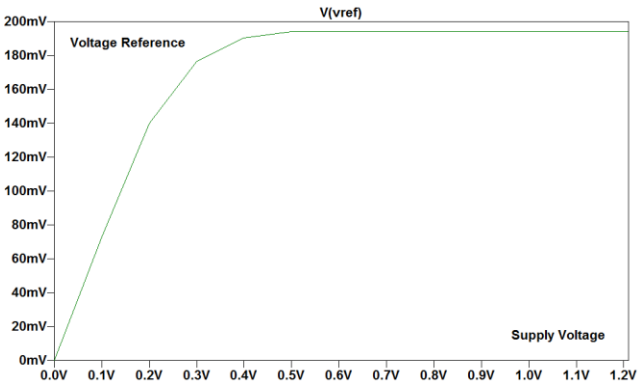


Figure 8: Voltage Reference vs Supply Voltage (Design I)

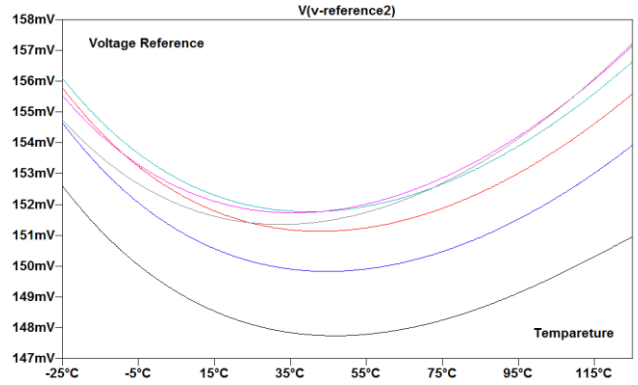


Figure 9: Voltage Reference vs Temperature (Design II)

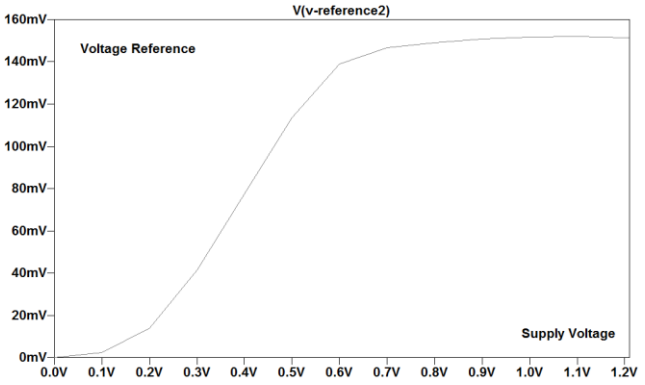


Figure 10: Voltage Reference vs Supply Voltage (Design II)

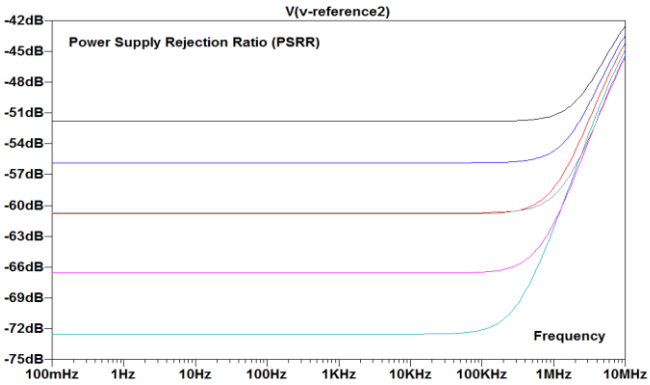


Figure 11: Power Supply Rejection Ratio vs Frequency (Design II)

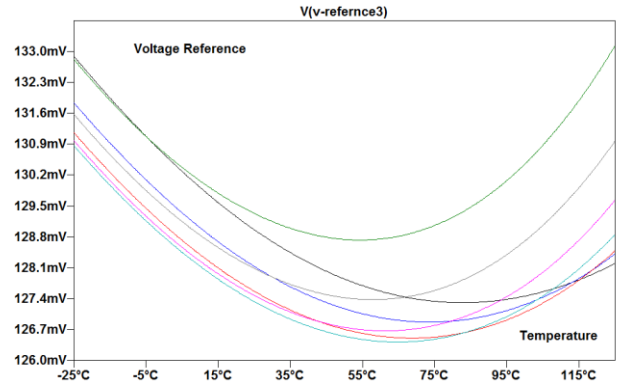


Figure 12: Voltage Reference vs Temperature (Design III)

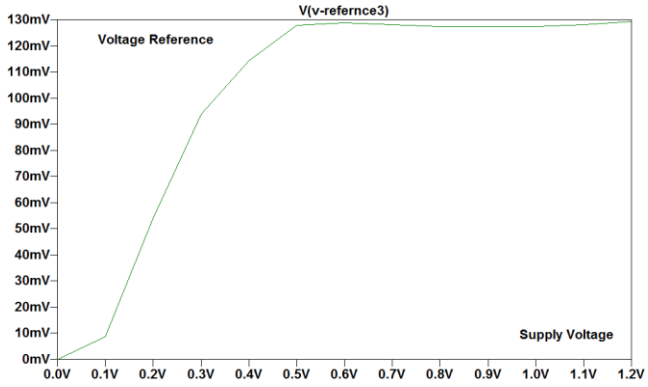


Figure 13: Voltage Reference vs Supply Voltage (Design III)

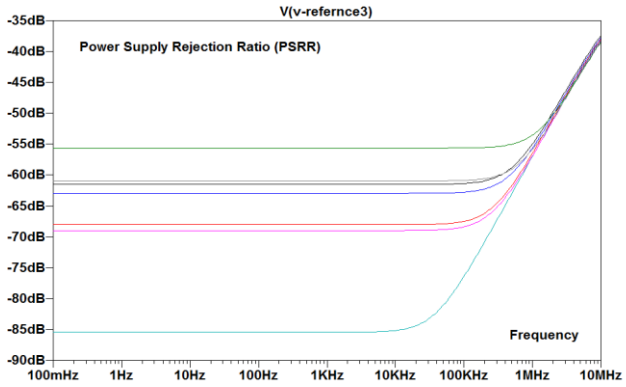


Figure 14: Power Supply Rejection Ratio vs Frequency (Design III)

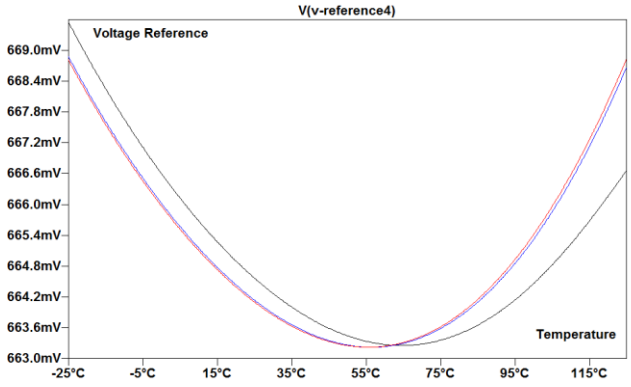


Figure 15: Voltage Reference vs Temperature (Design IV)

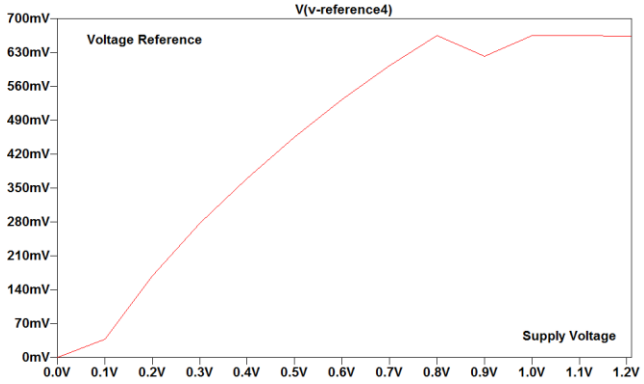


Figure 16: Voltage Reference vs Supply Voltage (Design IV)

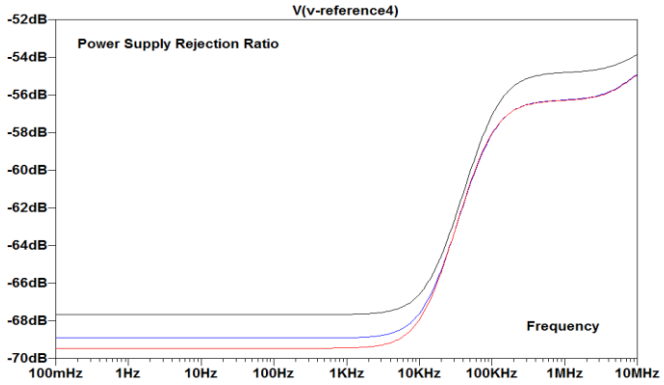


Figure 17: Power Supply Rejection Ratio vs Frequency (Design IV)

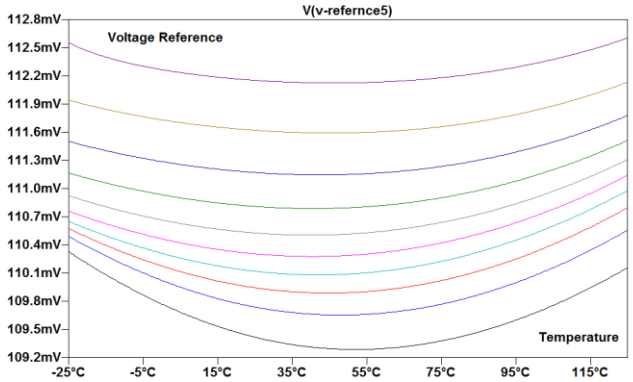


Figure 18: Voltage Reference vs Temperature (Design V)

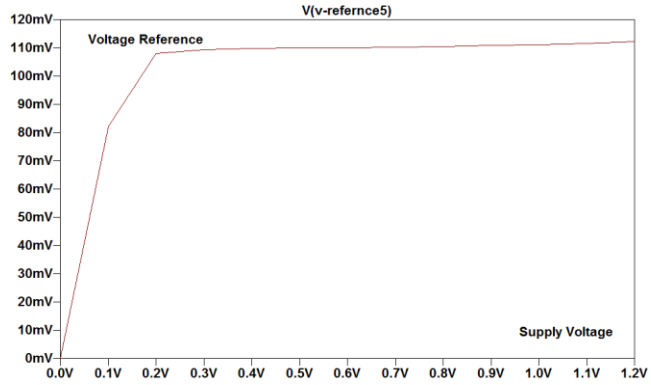


Figure 19: Voltage Reference vs Supply Voltage (Design V)

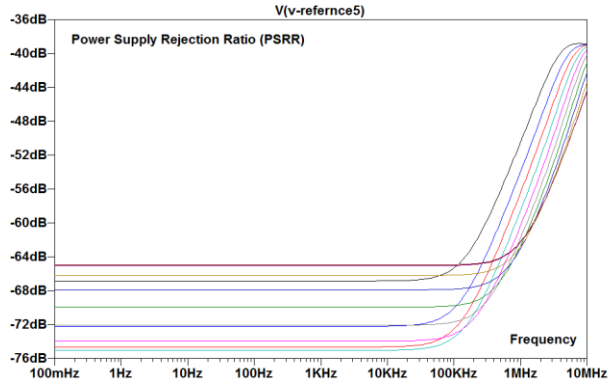


Figure 20: Power Supply Rejection Ratio vs Frequency (Design V)

Table1 : Performance Comparisons of Voltage Reference Circuits

Design		I [8]	II [9]	III [10]	IV [11]	V [12]
CMOS Technology		65nm	180nm	45nm	45nm	180nm
Minimum Supply Voltage [V]		0.8	0.9	0.8	0.6	0.5
Supply Voltage Range [V]		0.8 to 1.4	NA	NA	0.6 to1	0.5 to 2.2
Temperature Range [°c]		-40 to 125	-20 to 120	0 to 100	-25 to 85	-25 to 100
Temperature Coefficient [ppm/°c]		5.6	NA	14.62	19	11.4
Voltage Reference [mV]		428	221	310	193.20	211.46
Power Supply Rejection Ratio[dB]	10HZ	- 87	NA	NA	NA	- 61.5
	100KHZ	- 87	NA	NA	NA	NA
	1MHZ	- 75	NA	NA	- 52	NA
	10MHZ	NA	NA	NA	NA	NA
Gate Area [µm²]		468	3297.5	35	507.525	76.58
Results below are obtained using 65nm process						
Temperature Range [°c]		-25 to 125	-25 to 125	-25 to 125	-25 to 125	-25 to 125
Temperature Coefficient [ppm/ °c]		3.42	76	39	5.68	5.4
Minimum Supply Voltage [V]		0.6	0.75	0.6	1.1	0.3
Supply Voltage Range [V]		0.6 to 1.21	0.75 to 1.21	0.6 to 1.21	1.1 to 1.21	0.3 to 1.21
Voltage Reference @ Minimum Supply Voltage [mV]		196.5	122	132.9	668.8	110.3
Voltage Reference @ Minimum Supply Voltage + 10% [mV]		196.5	123.3	132.2	669.5	110.39
Power Supply Rejection Ratio [dB]	10HZ	-131.9	-50.7	-61.5	-65.4	-66.8
	100KHZ	-111.48	-50.7	-64.5	-66.28	-66.8
	1MHZ	-81	-50.7	-55.1	-65.56	-50
	10MHZ	-65	-43.5	-38	-57.3	-39
Transistor Count		16	11	11	19	17
Gate Area [µm²]		469.6	59.23	43.14	50.96	25.034

V. CONCLUSION

Selected voltage reference circuits reported in the literature are ported to the same process technology, namely 65nm CMOS process. This allowed us to compare performance parameters as a result of innovative circuit design techniques rather than process advancement. Design I yielded the best overall performance parameters (3.42 ppm/ºc, 0.6V minimum supply, -81dB PSRR @ 1MHz). However, this design is the largest in silicon area. Design V is very promising with the smallest silicon area and the lowest supply voltage.

This work will continue to generate comparisons for other types of voltage reference circuits (BJT based) in the same 65nm process so one can draw conclusions among all designs.

REFERENCES

[1] 65nm BSIM4 model card for bulk CMOS: V0.0, http://ptm.asu.edu/latest.html, accessed on 10.12.2020, .

[2] D. F. Hilbiber, "A New Semiconductor Voltage Standard," in International Solid-State Circuits Conference ISSCC, 1964.

[3] R. J. WIDLAR, “New Developments in IC Voltage Regulators,” IEEE JOURNALOF SOLID-STATE CIRCUITS, 1 2 1971.

[4] K. E. KUIJK, "A Precision Reference Voltage Source," IEEE JOURNAL OF SOLID-STATE CIRCUITS, 3 6 1973.

[5] A. P. BROKAW, "A Simple Three-Terminal IC Bandgap Reference," IEEE JOURNAL OF SOLID-STATECIRCUITS, 6 12 1974.

[6] B. M. J. K. a. M. J. Harry Neuteboom, “A DSP-Based Hearing Instrument IC,” IEEE JOURNAL OF SOLID-STATE CIRCUITS,, 11 11 1997.

[7] H. S. A. U. T. M. Hironori Banba, "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," IEEE JOURNAL OF SOLID-STATE CIRCUITS, , 5 5 1999.

[8] W. S. a. J. S. Jize Jiang, "A 5.6 ppm/°C Temperature Coefficient,87-dB PSRR, Sub-1 V Voltage Reference in 65-nm CMOS Exploiting the ZeroTemperature-Coefficient Ponit," IEEE JOURNAL OF SOLID-STATE CIRCUITS , 2016.

[9] H. L. Po-Hsuan Huang, "A Simple Subthreshold CMOS Voltage Reference Circuit With Channel-Length Modulation Compensation," IEEE J. Solid-State Circuits, 9 9 2006.

[10] S. S. ,. K. S. Rohit Singh, "Design and Optimization of a Low Power Voltage Reference Generator Circuit in 45nm CMOS Technology," INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT, pp. 298-303, January 2014.

[11] A. K. S. S. S. Rohit Singh, "Feedback Driven Temperature and Supply Compensated Sub-0.5 V Reference Generator," in 2014 9th International Conference on Industrial and Information Systems (ICIIS), 2014.

[12] R. Z. Q. S. a. H. Z. Yuwei Wang, "A 0.5 V, 650 pW, 0.031%/V Line Regulation Subthreshold Voltage Reference," in ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC), Dresden, Germany, 2018.

[13] R. S. MONIKA GUPTA, "Design of Ultra Low Power, Temperature Independent Resistorless Bandgap Reference Circuit in 0.18um CMOS," in 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), 2018.

Trapping Luminescent Nanoparticles: A Source of Low-Loss Optical Communications Photons

Jamal Ehtaiba
Deptartment of Electrical Technologies
The Higher Institute of Science and Technology
Misrata, Libya
ehtaiba_j@yahoo.co.uk

***Abstract*—Isolating a single ion to generate a nonclassical light is extremely difficult to accomplish, but synthesizing luminescent-ion-contained nanocrystals is a constantly evolving technique. Therefore, we here present an optical trapping technique for capturing a single nanocrystal with $1550\,\mathrm{nm}$ wavelength photon emission. The method uses a single-mode optical fiber probe with a metallic nanoantenna at the tip and a wave splitter at the back. Additionally, due to its ease of implementation and use, this trapping method can assist in developing single-photon sources for quantum fibre-optic communications.**

***Index Terms*—Photon-Emission, Single-Photon, Nanoantenna, Optical-Trapping, Quantum-Light, Optical-Communications**

I. Introduction

Although a broadband of optical spectrum is available, optimum quantum transmission through fiber-optic channel occurs only at the low-loss wavelength window, $1.55\,\mu\mathrm{m}$. Accordingly, tremendous investigations have been devoted to the search for clean (antibunched) light sources to replace conventional lasers and LEDs [1–4]. Information transmission and data exchange are more secure with quantum light emitters than chaotic light communication systems [5, 6]. Examples of quantum light sources include single atoms/ions [7–10], quantum dots [11–14], and vacancy centers [15, 16]. Proper synthesis of lanthanide ions of desirable photon emission wavelength with nanosize host materials can also be of attracting interest. For instance, nanoparticles (NPs), or solid-state nanocrystals (NCs) doped with lanthanide ions have been demonstrated as an excellent source of photons in the visible and near-infrared bands [17–21]. While single atoms and quantum dots are challenging to isolate/fabricate, NCs are handier to produce and more reliable [22–27]. The type of NCs, which we have used in our experimental work, is the hexagonal-phase sodium yttrium fluoride crystals co-doped with ytterbium and erbium ions, $\beta\text{-}\mathrm{NaYF_4} : \mathrm{Yb^{3+}/Er^{3+}}$.

Both up and downconversion processes can occur within the nanocrystal lattice structure when it is excited with a $980\,\mathrm{nm}$ laser beam. As a result, the upconversion process yields emission spectra in the visible band, while the downconversion process produces a peak at $1550\,\mathrm{nm}$ wavelength (see Fig. 1), the wavelength at which the lowest light attenuation occurs over fiber-optic routes. However, the essence of these conversion processes is in the kind of light produced from the single nanocrystal. Certainly, light emission from a single erbium-ion would be nonclassical quantum light, but as the number of emitters increases, chaos takes over.

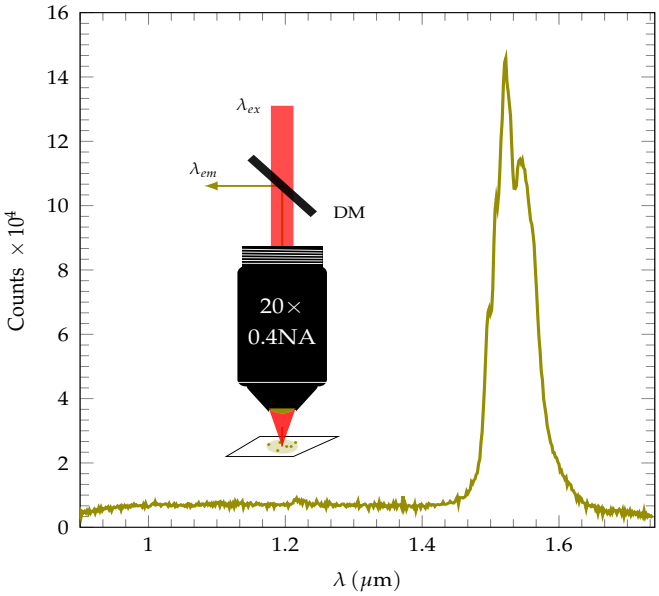


Fig. 1. Emission peak at $1550\,\mathrm{nm}$ from the Erbium ions contained in the $\mathrm{NaYF_4}$ NCs. The inset partially illustrates the optical setup. λ_{ex} represents the laser excitation wavelength ($\lambda_{ex} = 980\,\mathrm{nm}$), λ_{em} denotes emitted light, and DM represents a $1200\,\mathrm{nm}$ short-pass dichroic mirror.

Of particular note, planting a few Erbium ions in a single NC will more likely have a weak photon emission signal in addition to the long transition lifetime, and therefore both optical signal and emission rate enhancements are of important requirements. A variety of techniques have been demonstrated for this purpose utilizing plasmonic and/or photonic bandgap structures [28, 29].

As nanoholes in metal films can be used to provide effective optical trapping results in conventional optical trapping systems that use microscopes, the nanoaperture optical fiber tweezer (NAFT) is much easier to use without the need for complicated optics. The NAFT is a modern trapping technique composed of a pumping laser and a plasmonic nanoantenna made of a $100\,\mathrm{nm}$ thick gold film placed at the tip of a single-mode optical fiber (SMF). The nanoantenna is the key component of NAFT since it contains a nanoscale

gap capable of confining a high field intensity and setting up the optical trapping gradient forces. More details on the nanoantenna (nanoaperture) and how it is fabricated and placed at the fiber tip are provided in [30]. Here, we demonstrate optical trapping results of an Erbium-doped single nanocrystal (β-$NaYF_4 : Yb^{3+}/Er^{3+}$) of 30 nm average size using NAFT.

II. Photon Emission and NC Trapping Experiments

A. Measuring the 1550 nm Optical Emission of the NCs

Fig. 1 shows the 1550 nm emission spectrum of the NCs under infra-red laser beam excitation. We obtained the emission spectrum using the setup arrangement shown partially in Fig. 1 inset, where a sample of nanocrystals lies on 100 nm thick gold film with a glass substrate. The sample was a cluster (about 1 mg) of the NPs embedded inside a very thin epoxy shield to avoid oxidation and interaction with the surrounding environment. In addition, the purpose of the gold film was to enhance the optical emission from the NCs.

The sample was excited with a collimated and focused 980 nm laser beam through an objective lens ($20\times$, 0.4NA). The laser output power is 20 mW and the incident power density on the sample is approximately $2.8\,\mathrm{mW}/\mu\mathrm{m}^2$. Light emitted from the NCs was collected by the same objective lens and partially passed through a dual mirror that allows easy extraction to the 1550 nm signal from the total collected field.

B. Optical Trapping of the NC

This part illustrates the optical trapping process for the luminescent NCs. Fig. 2 inset shows a schematic diagram of the trapping setup. The setup consists of; NPs suspension medium (Hexane, a nonpolar liquid), the NAFT, a 980/1550 nm optical fiber wavelength division multiplexer (WDM), and an APD (Avalanche Photodetector), which is not shown in the figure. The fiber WDM has three ports; the red line represents the 980 nm port; the blue line represents the 1550 nm port; and the black line represents the common port. The WDM was designed to split or combine waves at the two different wavelengths. The excitation laser beam (about 20 mW at 980 nm) is inserted into the red arm of the WDM through an isolator and a polarizer (both are not shown in the figure). The function of the polarizer is to enhance field intensity within the nanoantenna gap and hence increase trapping force. However, nanoparticle trapping events can be identified by changes in the electric field intensity resulting from changes in the refractive index within the nanoantenna gap structure. Ultimately, it was of course necessary for the nano-antenna design to have a maximum transmittance at 1550 nm wavelength that will allow efficient coupling of any light emission from the nano-particles into the fiber channel.

We have performed the optical trapping experiments in two repeatable procedures. First, we used pure hexane (without NPs) as a trapping medium, and in the second case, we incorporated 0.1 g of NCs to the pure hexane to form a trapping medium of $0.1\%(w/v)$ concentration. A time record of the trapping signal for the case of pure hexane is shown in

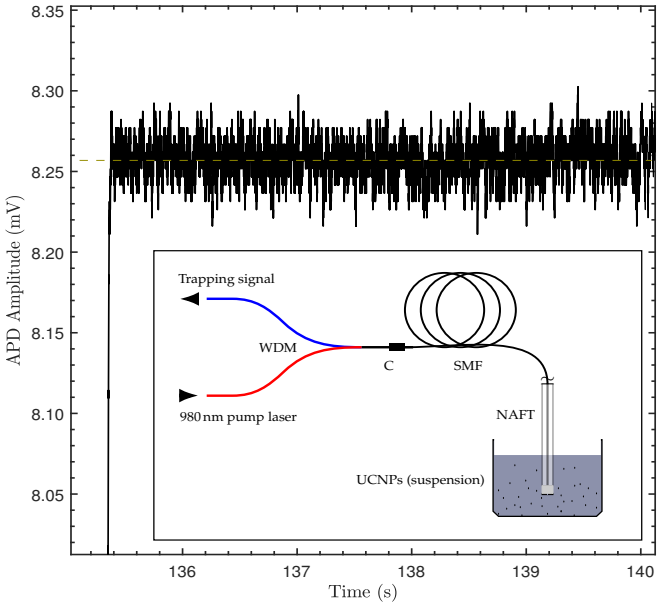


Fig. 2. Optical trapping signal trace for pure hexane (without NCs). The dashed line represents the signal fluctuation average, and the inset shows a schematic for the trapping experiment setup.

Fig. 2, where the signal start-time (around 135.35 s) represents the moment the laser was switched on. It is obvious from the signal trace that there is no change in the trapping signal fluctuations (and in the signal average value) even when the laser is being switched on and off multiple times.

This response indicates that the electric field intensity within the nanoantenna gab hasn't changed or, more predictively, there has been nothing captured by the nanoantenna, the NAFT.

In Fig. 3, we show the optical trapping tracing signal for the case when we had NPs suspended in hexane solvent ($0.1\%(w/v)$ concentration), shown as UCNPs in Fig. 2 inset. To separate NPs agglomeration and uniformly distribute them over the entire hexane volume, we sonicated the mixture for about half an hour just before use. The sonication process enhances the optical trapping of individual nanoobjects and helps achieve our goal to trap a single luminescent NP. However, the optical trapping signal shows a jump in the signal level a few hundred milliseconds after the laser has turned on. This event reveals that a nanoparticle has pulled into the gap of the nanoantenna influenced by the optical field trapping forces. Hence the signal fluctuation and its average have both increased. As Fig. 2 and Fig. 3 illustrate, two differences in the trapping signals are visible. First, the change in the signal average and its range of amplitude fluctuations around the average. Notably, the signal average in Fig. 2 is constant over time, and the fluctuations around the average are small. On the other hand, the optical trapping signal of case 2 (see Fig. 3) shows a remarkable change in the average value and amplitude fluctuation range. And this can be attributed to the electric field intensity fluctuations due to disorders in the refractive index caused by a moving nanoobject entering the nanoantenna

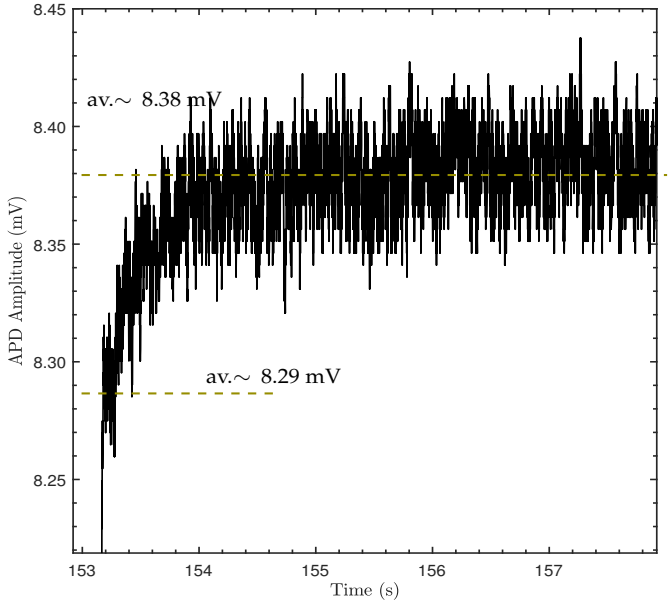


Fig. 3. Optical trapping signal trace for NPs contained hexane. The dashed line represents the signal fluctuation average.

gap structure. In the end, the signal level jump in case 2 can be explained by the nanoantenna capturing a nanoparticle, trapping.

Although the nanoantenna has an optical transmission peak at the wavelength of $1550\,\mathrm{nm}$, we were unable to observe any light emitting from the nanoparticle when using a spectrometer, and this could be due to three factors: no emission from the nanoparticle; weak radiation of light; or, thermal noise.

III. Conclusion

We have experimentally demonstrated the ability to trap nanocrystals that feature photon emission at low-loss wavelengths needed in optical communications. The optical trapping technique utilized in the trapping process combines an optical fiber tweezer and the WDM. Detection of light emission from trapped nanocrystals was difficult mainly due to temperature, but further investigations can be done in future, expectedly with better results, especially when the optical trapping system combines with a superconducting nanowire single-photon detector. Overall, we have successfully trapped a single light emitter and have coupled antenna light transmission into the optical fiber channel. Eventually, optical trapping of luminescent nanoparticles allows the potential of developing single-photon sources, which play an essential role in fiber optic communications and quantum data transmission.

Acknowledgment

The author acknowledges technical support from the University of Victoria, B.C., Canada. Thanks to R. Gordon.

References

[1] M. Di Giancamillo, P. Biagioni, V. Sorianello, and E. Prati, "Design of erbium doped silicon nanocavities for single photon applications," in *EPJ Web of Conferences*, vol. 255. EDP Sciences, 2021.

[2] X. He, H. Htoon, S. Doorn, W. Pernice, F. Pyatkov, R. Krupke, A. Jeantet, Y. Chassagneux, and C. Voisin, "Carbon nanotubes as emerging quantum-light sources," *Nature materials*, vol. 17, no. 8, pp. 663–670, 2018.

[3] S. Rodt, S. Reitzenstein, and T. Heindel, "Deterministically fabricated solid-state quantum-light sources," *Journal of Physics: Condensed Matter*, vol. 32, no. 15, p. 153003, 2020.

[4] S. Strauf, "Towards efficient quantum sources," *Nature Photonics*, vol. 4, no. 3, pp. 132–134, 2010.

[5] P. Michler, A. Kiraz, C. Becher, W. Schoenfeld, P. Petroff, L. Zhang, E. Hu, and A. Imamoglu, "A quantum dot single-photon turnstile device," *science*, vol. 290, no. 5500, pp. 2282–2285, 2000.

[6] P. Yao, V. Manga Rao, and S. Hughes, "On-chip single photon sources using planar photonic crystals and single quantum dots," *Laser & Photonics Reviews*, vol. 4, no. 4, pp. 499–516, 2010.

[7] H. J. Kimble, M. Dagenais, and L. Mandel, "Photon antibunching in resonance fluorescence," *Physical Review Letters*, vol. 39, no. 11, p. 691, 1977.

[8] P. Pinkse, T. Fischer, P. Maunz, and G. Rempe, "Trapping an atom with single photons," *Nature*, vol. 404, no. 6776, pp. 365–368, 2000.

[9] J. McKeever, A. Boca, A. Boozer, R. Miller, J. Buck, A. Kuzmich, and H. Kimble, "Deterministic generation of single photons from one atom trapped in a cavity," *Science*, vol. 303, no. 5666, pp. 1992–1994, 2004.

[10] S. Wolf, S. Richter, J. von Zanthier, and F. Schmidt-Kaler, "Light of two atoms in free space: bunching or antibunching?" *Physical review letters*, vol. 124, no. 6, p. 063603, 2020.

[11] S. Fasel, O. Alibart, S. Tanzilli, P. Baldi, A. Beveratos, N. Gisin, and H. Zbinden, "High-quality asynchronous heralded single-photon source at telecom wavelength," *New Journal of Physics*, vol. 6, no. 1, p. 163, 2004.

[12] K. Rivoire, S. Buckley, A. Majumdar, H. Kim, P. Petroff, and J. Vučković, "Fast quantum dot single photon source triggered at telecommunications wavelength," *Applied Physics Letters*, vol. 98, no. 8, p. 083105, 2011.

[13] Z.-S. Chen, B. Ma, X.-J. Shang, H.-Q. Ni, J.-L. Wang, and Z.-C. Niu, "Bright single-photon source at 1.3 μm based on inas bilayer quantum dot in micropillar," *Nanoscale research letters*, vol. 12, no. 1, pp. 1–6, 2017.

[14] P. Holewa, A. Sakanas, U. M. Gür, P. Mrowiński, A. Huck, B.-Y. Wang, A. Musiał, K. Yvind, N. Gregersen, M. Syperek *et al.*, "Bright quantum dot single-photon emitters at telecom bands heterogeneously integrated on si," *arXiv preprint arXiv:2104.07589*, 2021.

[15] E. Neu, M. Agio, and C. Becher, "Photophysics of single silicon vacancy centers in diamond: implications for single photon emission," *Optics express*, vol. 20, no. 18, pp. 19 956–19 971, 2012.

[16] S. I. Bogdanov, M. Y. Shalaginov, A. S. Lagutchev, C.-C. Chiang, D. Shah, A. S. Baburin, I. A. Ryzhikov, I. A. Rodionov, A. V. Kildishev, A. Boltasseva *et al.*, "Ultrabright room-temperature sub-nanosecond emission from single nitrogen-vacancy centers coupled to nanopatch antennas," *Nano letters*, vol. 18, no. 8, pp. 4837–4844, 2018.

[17] Z. Li and Y. Zhang, "An efficient and user-friendly method for the synthesis of hexagonal-phase nayf4: Yb, er/tm nanocrystals with controllable shape and upconversion fluorescence," *Nanotechnology*, vol. 19, no. 34, p. 345606, 2008.

[18] M. Ding, C. Lu, L. Cao, J. Song, Y. Ni, and Z. Xu, "Facile synthesis of β-nayf 4: Ln 3+(ln= eu, tb, yb/er, yb/tm) microcrystals with down- and up-conversion luminescence," *Journal of Materials Science*, vol. 48, no. 14, pp. 4989–4998, 2013.

[19] X. Yin, H. Wang, Y. Tian, M. Xing, Y. Fu, and X. Luo, "Three primary color emissions from single multilayered nanocrystals," *Nanoscale*, vol. 10, no. 20, pp. 9673–9678, 2018.

[20] B. Casabone, C. Deshmukh, S. Liu, D. Serrano, A. Ferrier, T. Hümmer, P. Goldner, D. Hunger, and H. de Riedmatten, "Dynamic control of purcell enhanced emission of erbium ions in nanoparticles," *Nature communications*, vol. 12, no. 1, pp. 1–7, 2021.

[21] M. Celebrano, L. Ghirardini, M. Finazzi, Y. Shimizu, Y. Tu, K. Inoue, Y. Nagai, T. Shinada, Y. Chiba, A. Abdelghafar *et al.*, "1.54 μm photoluminescence from er: O x centers at extremely low concentration in silicon at 300 k," *Optics letters*, vol. 42, no. 17, pp. 3311–3314, 2017.

[22] A. D. Ostrowski, E. M. Chan, D. J. Gargas, E. M. Katz, G. Han, P. J. Schuck, D. J. Milliron, and B. E. Cohen, "Controlled synthesis and single-particle imaging of bright, sub-10 nm lanthanide-doped upconverting nanocrystals," *ACS nano*, vol. 6, no. 3, pp. 2686–2692, 2012.

[23] D. J. Gargas, E. M. Chan, A. D. Ostrowski, S. Aloni, M. V. P. Altoe, E. S. Barnard, B. Sanii, J. J. Urban, D. J. Milliron, B. E. Cohen *et al.*, "Engineering bright sub-10-nm upconverting nanocrystals for single-molecule imaging," *Nature nanotechnology*, vol. 9, no. 4, pp. 300–305, 2014.

[24] K. Huang, N. M. Idris, and Y. Zhang, "Engineering of lanthanide-doped upconversion nanoparticles for optical encoding," *Small*, vol. 12, no. 7, pp. 836–852, 2016.

[25] J. Hesse, D. T. Klier, M. Sgarzi, A. Nsubuga, C. Bauer, J. Grenzer, R. Hübner, M. Wislicenus, T. Joshi, M. U. Kumke *et al.*, "Rapid synthesis of sub-10 nm hexagonal nayf4-based upconverting nanoparticles using therminol® 66," *ChemistryOpen*, vol. 7, no. 2, p. 159, 2018.

[26] S. Wen, J. Zhou, K. Zheng, A. Bednarkiewicz, X. Liu, and D. Jin, "Advances in highly doped upconversion nanoparticles," *Nature communications*, vol. 9, no. 1, pp. 1–12, 2018.

[27] X. Zhu, J. Zhang, J. Liu, and Y. Zhang, "Recent progress of rare-earth doped upconversion nanoparticles: synthesis, optimization, and applications," *Advanced Science*, vol. 6, no. 22, p. 1901358, 2019.

[28] J. Liao, Z. Yang, H. Wu, D. Yan, J. Qiu, Z. Song, Y. Yang, D. Zhou, and Z. Yin, "Enhancement of the up-conversion luminescence of yb 3+/er 3+ or yb 3+/tm 3+ co-doped nayf4 nanoparticles by photonic crystals," *Journal of Materials Chemistry C*, vol. 1, no. 40, pp. 6541–6546, 2013.

[29] J. Liao, Z. Yang, S. Lai, B. Shao, J. Li, J. Qiu, Z. Song, and Y. Yang, "Upconversion emission enhancement of nayf4: Yb, er nanoparticles by coupling silver nanoparticle plasmons and photonic crystal effects," *The Journal of Physical Chemistry C*, vol. 118, no. 31, pp. 17 992–17 999, 2014.

[30] J. M. Ehtaiba and R. Gordon, "Template-stripped nanoaperture tweezer integrated with optical fiber," *Optics express*, vol. 26, no. 8, pp. 9607–9613, 2018.

Performance Analysis of the High-Speed WDM-FSO-MIMO System Under Different Weather Conditions

Ahmed Mohammed Abdel Aal
Faculty of Electrical and Electronic Engineer
University of Tripoli
Tripoli, Libya
A.Abdul-all@uot.edu.ly

Khaled Elgdamsi
Faculty of Electrical and Electronic Engineer
University of Tripoli
Tripoli, Libya
K.elgdamsi@uot.edu.ly

***Abstract* — Free-space optics (FSO) is a data relaying technology, which requires a direct line of sight between the transmitter and the receiver units for reliable transmission. Wavelength Division Multiplexing (WDM) is a technology that multiplexes numerous carrier signals onto single fiber using nonidentical wavelengths and enables the efficiency of bandwidth and expanded data rate. Multiple Input Multiple Output (MIMO) is implemented to improve the quality and performance of free space optical communication in various atmospheric conditions. In this paper, a WDM-based FSO communication system is being implemented under MIMO concept i.e., 1 × 1, 2 × 2, 4 × 4 Various factors like BER and Quality Factor are analyzed for the WDM-based FSO communication with MIMO using the OptiSystem v 7.0 for under different atmospheric conditions. It is also evident in this paper that the transmit power of 15 dBm and 4 × 4 FSO showed the best performance and highest correlation distance compared to the transmit power of 10dBm and other beams. Moreover, WDM-FSO MIMO systems with 20 cm receiving aperture better than 10 cm. In addition, systems with a data rate of 2.5 $Gbps$ provide better performance than 5 $Gbps$.**

Keywords— Free Space Optics (FSO); bit error rate and Q-factor; multiple input multiple output; OptiSystem; wavelength division multiplexing; Attenuation.

I. INTRODUCTION

Free space optical (FSO) communication now a days is one of the major topics in the world of wireless and optical communications. This type of wireless optical communication technology uses highly narrow beam to transmit data from one point to the other one. The clear line of sight between both transmits and receive terminals is essential to establish a seamless communication [1]. FSO communication links have many merits such as high modulation bandwidth, high data transmission rates, low cost, and easy installation process. That is optical communication technology in which data is transmitted by propagation of light in free space allowing optical connectivity. There is no requirement of the optical fiber cable. Working of FSO is similar to OFC (optical fiber cable) networks but the only difference is that the optical beams are sent through free air instead of OFC cores that is glass fiber. FSO system consists of an optical transceiver at both ends to provide full duplex (bidirectional) capability. FSO communication is not a new technology. It has been in existence from 8th century but now is more evolved, where data, voice, and video communication are achieved with maximum 10Gbps of data rate by full duplex (bidirectional) connectivity [2].

free space optics have many advantages over any conventional system like huge bandwidth which help us to transmit large number of data, narrow beam divergences which help us to transmit the signal to long distances with less misalignment [3]. FSO refers to transmission in unguided propagation media through the use of optical carriers, i.e., visible, infrared (IR) and ultraviolet (UV) band. As shown in electromagnetic spectrum illustrated in Figure 1 [4].

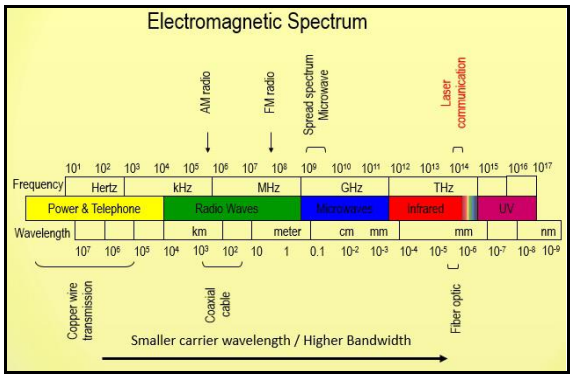


Figure 1: The electromagnetic spectrum.

The 1550 nm band is well suited for free space transmission due to its low attenuation, as well as the proliferation of high-quality transmitter and detector components. Components include very high-speed semiconductor laser technology suitable for Wavelength Division Multiplex (WDM) operation [5]. The FSO transmitter, located on the first building and connected to a Local Area Network (LAN) situated in that building, transmits the optically modulated signal through the free space. The FSO receiver, located on the second building and connected also to a LAN situated in that building, receives the transmitted signal as shown in Figure 2. [6].

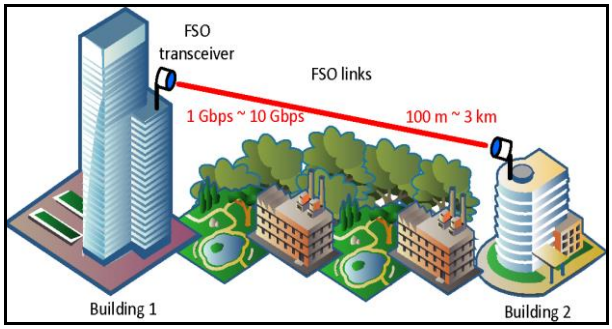


Figure 2: Point-to-point FSO link between two buildings.

The transfer rate of a physical optical link could be increased with using of multichannel technology WDM. That could be used to achieve higher transmission rates and keep same physical components of FSO link. The FSO link can work at a wavelength of 1550 nm which is also in transmission window with a low spectral attenuation.

In order to increase the data rate in free space and to reduce the attenuation due to weather conditions, WDM-FSO-MIMO performance will be used and analyzed in this paper

The WDM innovation represents a revolution within the optical communication. WDM can grow the capacity incrementally, and that WDM provides great simplicity and flexibility in the network. These technologies, on the other hand, enable optical multi- and demultiplexing because individual signals have different light wavelengths and can be separated easily by wavelength-selective optical elements. This may enable us to construct WDM networks in which node functionality is supported by optical technologies without electrical mux/demux [7].

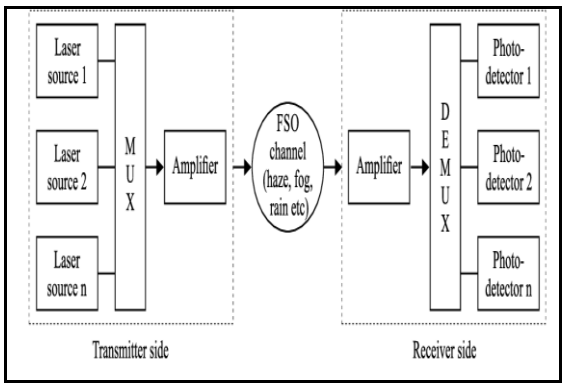


Figure 3: Block diagram of WDM-FSO system.

FSO systems are categorized into two types, FSO system with single beam and FSO system with multiple-beams. The systems categories have positive and negative attributes. Multi-beam beam FSO system is shown in Figure 4, where the goal is to transmit data through one beam only. The main drawback of a single beam FSO system is that the beam will be scattered while travelling in free space, which is due to atmospheric turbulence arising from big seized raindrops or haze. Therefore, the probability of beam reaching the receiver is very low in these situations, as a solution, multiple beams are used to increase the probability by assuring at least one of the multiple beams to reach the target [8].

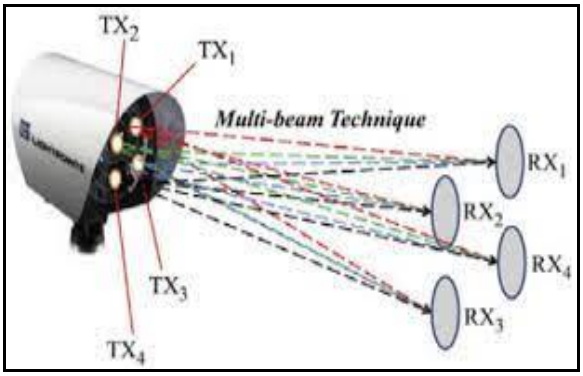


Figure 4: Multi-beam FSO system.

The multiple TX/RX link configuration can be seen in Figure 4. Two FSO terminals which each having a link head consists of multiple lenses of TXs and RXs. these lenses will produce and collect multiple laser beams along the optical path. the multiple beams which are the redundant signals are generated by the data splitter which splits the digital signal into 4 independent signals; each carries the same data input and leave the TXs as an independent beam, from figure Each TX will transmit 4 lasers beams to each of the 4 RXs because of the 4 data splitters used. In, total there would be 16 paths of laser beams/combinations of TX and RX to be analyzed [9].

II. WDM-FSO SYSTEM DESIGN CONSIDERATIONS (CHALLENGES)

It is important to note that the signal loss between the transmitter and the receiver varies randomly over time according to the instantaneous prevailing weather conditions. Hence, the design needs to factor this into account for reliable communication design. Well-known weather conditions that affect adversely the FSO communication is rain, fog and haze conditions.

A. Fog attenuation

Fog is a cloud of small pieces of water, smoke, ice, or integration between them it's founded near the earth's surface. Fog is the main reason that makes scattering of light to happen which reduces the vision. the fog attenuation is depended on the radius of the particle size. Kim and Kruse introduced a law to find the amount of attenuation depends only on the range of visibility, not on particle size. The visibility is defined as the distance at which a naked human eye can distinguish between white boundaries and the black boundaries and range of visibility is calculated at 550 nm wavelength that matches the highest intensity of the solar spectrum [10].

The attenuation for both Kim and Kruse model by depending on the range of visibility has been calculated according to equation (1) [11].

$$\sigma\left[\frac{1}{Km}\right] = \frac{3.91}{v}\left(\frac{\lambda}{550nm}\right)^{-q} \quad (1)$$

Where:

σ, denote to attenuation coefficient (in $\mathbf{km^{-1}}$).

v, denotes to the range of visibility (in km).

q, denotes to particle size distribution coefficient which based on visibility range.

λ, denotes to the wavelength of the transmitted signal (in nm).

These models use different equations for the calculation of particle size distribution which are as shown below:

(1) Kruse model: This model is widely used to calculate the FSO equipment link budget. This model corresponds to best results on the frequency 1550 nm rather than other frequencies suggested to be used in the FSO communication. The equation used in this model has the form shown below: The particle size distribution (q) for this model has the form [12]:

$$q = \begin{cases} 1.6 & if\ v > 50\ km \\ 1.3 & if\ 6km < v < 50\ km \\ 0.585\, v^{1/3} & if\ v < 6\ km \end{cases} \quad (2)$$

(2) Kim model: The evaluation of the parameter q was not collected in dense fog for the visibility lower than 6 km. So, for the visibility lower than 1 km the significance of the Kruse model was in doubt. Then the recent studies proposed another expression for the particle size distribution coefficient (q). this model was the Kim model, whose expression has the form shown below [11,12]:

$$q = \begin{cases} 1.6 & if\ v > 50km \\ 1.3 & if\ 6km < v < 50km \\ 0.16v + 0.34 & if\ 1km < v < 6km \\ v - 0.5 & if\ 0.5km < v < 1km \\ 0 & if\ v < 0.5\ km \end{cases} \quad (3)$$

where 6 $Km < V < 50\ Km$, 1 $Km < V < 6\ Km$ and 0 $Km < V < 1\ Km$ specify the visibility ranges for clear, hazy and foggy weather conditions, respectively.

B. Haze attenuation

Haze particles present in the air reduce the intensity of the light beam thus hindering the transmission. Attenuation caused by haze depends on the visibility level and is calculated by the famous Kim and Kruse model given by equation (1) [13].

According to the Beers-Lambert law, the relationship between the transmitted signal power (P_T) and the received signal power (P_R), in the presence of atmospheric attenuation, can be represented by the following exponential relation [14]:

$$P_R = P_T\ exp\ (-\sigma L) \quad (4)$$

Where, L is the link distance (distance between the transmitter and receiver) of point-to-point FSO link (in km) and σ is the atmospheric attenuation coefficient.

According to the definition of attenuation, the total atmospheric attenuation throughout the free space path due to fog and haze can be represented as follows [14]:

$$\alpha_{dB} = 10 \log \left(\frac{PT}{PR}\right) = 10 \log\ (exp\ (\sigma L)) \quad (5)$$

Hence, the atmospheric attenuation in (dB/km) can be given by:

$$\alpha\ (dB/Km) = \alpha\ (dB)/L = [10\ \log(exp(\sigma L))]\ /L \quad (6)$$

C. Rain attenuation

Rain scattering is formed due to the rainfall when the radius of raindrops ranges between (100 – 1000 μm) that is significantly greater than the wavelength of standard FSO systems. The value of the attenuation of the optical signal due to rain also depends on the number of rain droplets and size. The optical signal attenuation linearly raises when the rate of rain increases. However, the average of the rainfall rate increases when the raindrop sizes increase [10].

The attenuation of the wireless optical link which caused by rain is given by Carbonneau's models in equation (7).

$$\alpha_{rain}\ (dB\ /\ Km) = 1.07\ R^{0.67} \quad (7)$$

Where:

α_{rain}, denotes to attenuation of the rain in (dB/km).

R, denotes to rain Rate in (mm/h).

III. WDM-FSO MIMO Performance Evaluation

1) Bit Error Rate (BER)

In digital transmission, the BER value is determined as a ratio of the number of bit errors and the total number of transmitted bits during a studied time interval. BER is another basic qualitative parameter of FSO link. In this work, has defined BER as the estimation of equation (8) where, ne is the number of received error bits and NB is the number of all transmitted bits for a long period [9].

$$BER \approx \frac{ne}{NB} \quad (8)$$

2) Quality Factor

Q is defined for a digital transmission signal as a signal-to-noise ratio (SNR) at the receiver's decision circuit and this parameter use to illustrate and measure the link quality is expressed as [15]:

$$Q = \frac{\mu_1 - \mu_0}{\delta_1 + \delta_0} \quad (9)$$

where μ_1 and μ_0 is the mean voltage level of the 1 and 0 levels. δ_1 and δ_0 are the standard deviations of the noise distribution on the 1 and 0 levels.

The quality factor of the received signal is calculated under different weather conditions by changing the transmission distance between the transmitted and received stations. The quality factor is related to BER as in equation (10) [10].

$$BER = \frac{1}{2}\ efrc\ \left(\frac{Q}{\sqrt{2}}\right)$$

$$BER \approx \frac{exp\ \left(\frac{-Q^2}{2}\right)}{Q\ \sqrt{2\ \pi}} \quad (10)$$

Where:

BER, denotes to the bit error rate of the received signal.

Q, the quality factor of the received signal.

$efrc$ is the error function complementary.

IV. WDM-FSO MIMO System Design

WDM is used to enhance the communication range and the capacity by implementing number of channels in the transmitter and receiver sections. The design consists of additional components like forks, power combiners to achieve the required performance. The MIMO technique has been used to implement this design which improves the BER, Quality factor of the FSO communication System.

A Software tool namely OptiSystem 7.0 is used to prove this practically by analyzing the parameters like BER, Quality factor.

In the simulation design, Figure 5 illustrates the schematic diagram of the proposed system, WDM-FSO MIMO.

The schematic diagram explains three major parts: the transmitter, channel, and receiver.

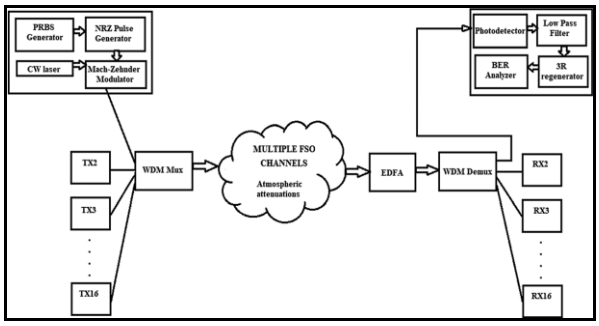


Figure 5: Schematic diagram of the proposed system, WDM-FSO MIMO.

The transmitter implemented by subsystems and WDM multiplexer (WDM Mux) and fork. each subsystem includes the following components as shown in Figure 6 : -

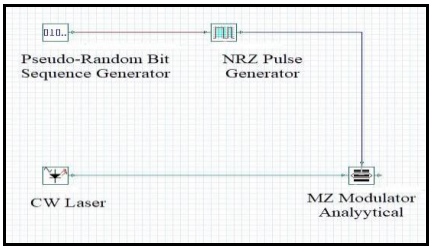


Figure 6: Transmitter subsystems.

- (CW Laser) Power source: The power source is represented by lasers diode that has a narrow beam width which provides a continuous wavelength of the laser signal. The laser signal is utilized as a carrier to modulate the input electrical signal.

- Pseudo-Random-Binary-sequence (PRBS) generator: This part responsible of generate random binary sequence. The PRBS generator produces binary sequence from zeros and ones

- Non-return to zero (NRZ) generator: This component is a pulse generator that is used for encoding the bit sequence that has been generated by PRBS generator.

- Mach-Zehnder modulator (intensity modulator): This modulator is utilized to modulate the electrical signal by using a light signal.

WDM multiplexer (WDM Mux) is the second part of the transmitter that is used to multiplex different optical signals wavelengths over one optical communication channel. The gain of the amplifier is set to 10 dB. The amplified signal is then directly sent to the receiver through FSO channel. Fork component is used to provide multiple beams of the laser from one source of the laser. Every beam of multiple lasers that have been produced by fork is in the same value with the input laser source from the previous component and then send in the FSO channel [10].

The FSO channel comprises of (4× 4), (2× 2), (1× 1) MIMO technique having various attenuations for different weather conditions. The receiver part consists of the following components:

- Power combiner is utilized to integrate transmitted optical signals and then injected in WDM DE multiplexer then reverses of transmission operation begin.

- WDM DE multiplexer (WDM Demux) is used to DE multiplex the received optical signals. In other words, it reverses the function of the WDM multiplexer.

- Optical receiver: After the operation of DE multiplexer, the photo signal enters the optical receiver. The optical receiver as shown in Figure 7.

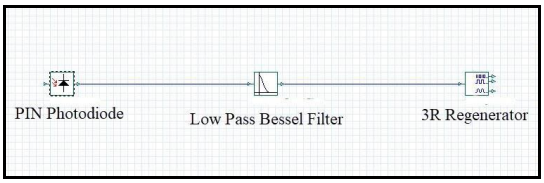


Figure 7: Optical receivers.

which consists of PIN photodetector that is used to detect and convert the photo signal into an electrical signal. The detected electrical signal is applied to a low pass Bessel filter. The low pass Bessel filter with a cut-off frequency of ($0.75\times symbol\ rate$) which is the second part of the optical receiver is used to split the modulated information from the high carrier frequency. The used low pass Bessel filer also minimizes the noise that is produced in the detection process. 3R regenerator is the third part of the optical receiver that is used to produce original sequence of bit and modulated electrical signal to be applied to BER analysis [10].

The BER analyzer is utilized to compute the BER of the received signal and quality factor. The measurements are taken from the first channel at a frequency of 193.5 THz.

Simulation layout using software OptiSystem 7.0 shown in Figure 8 for single input single output (SISO) and Figure 9 for MIMO.

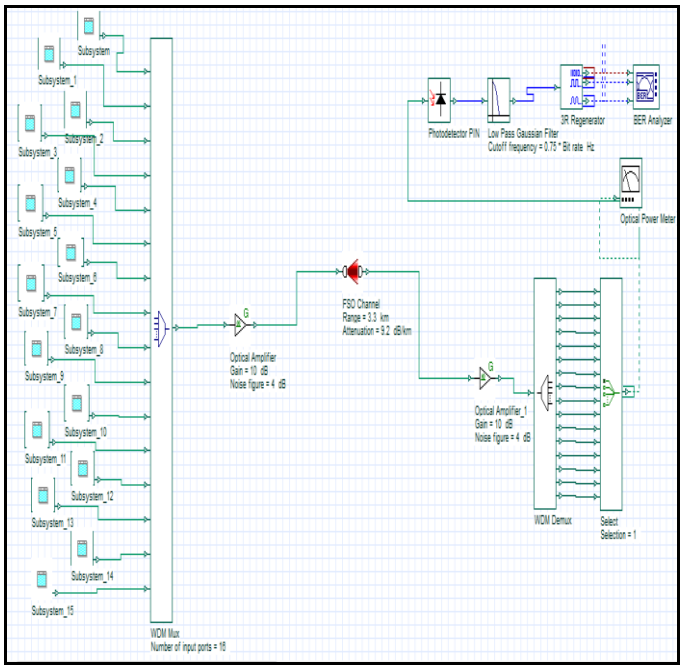
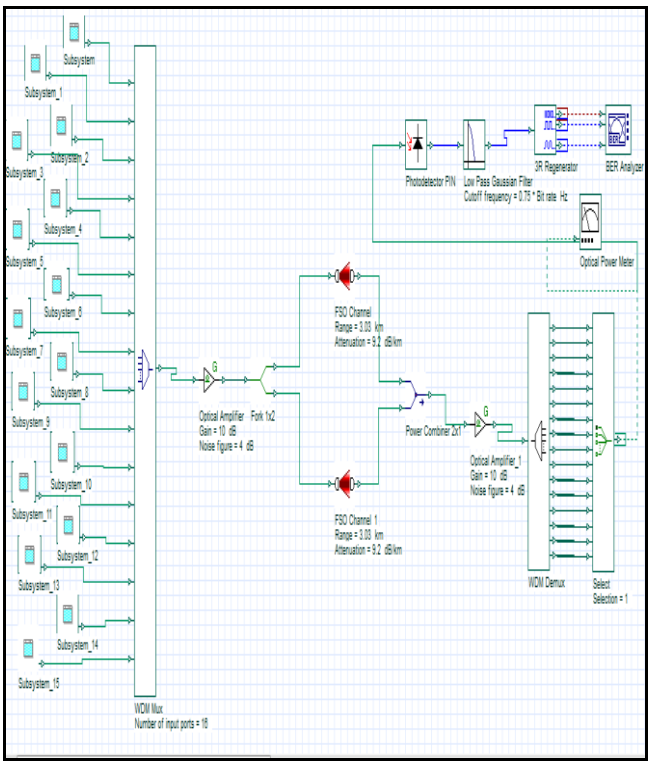
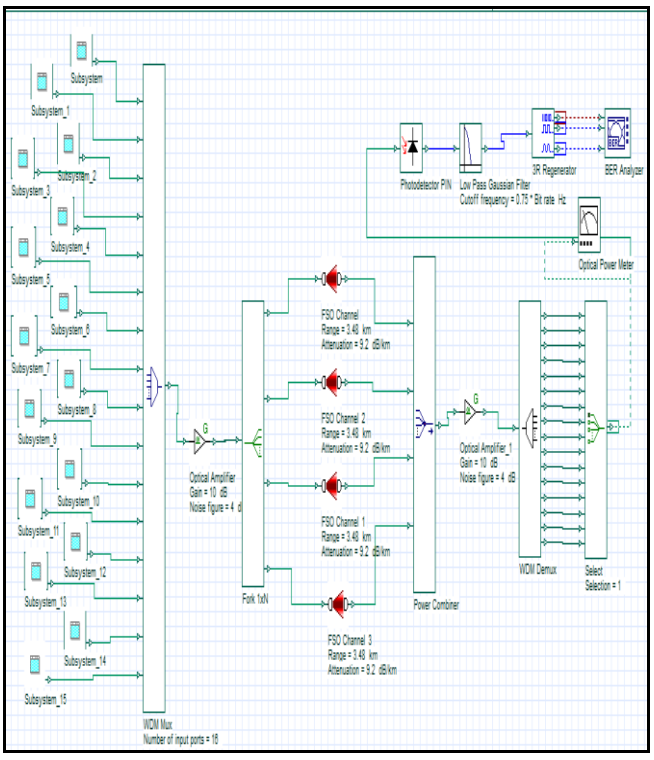


Figure 8: Simulation layout of **1-TX/1-RX (SISO)** system in WDM-FSO links.



(a)



(b)

Figure 9: (a) Simulation layout of **2-TX/2-RX** system in WDM-FSO links. **(b)** Simulation layout of **4-TX/4-RX** system in WDM-FSO links.

The link parameters with their symbols and corresponding values are presented in Table 1.

Table 1: WDM-FSO MIMO link parameters.

Parameter	Symbol	Value
Transmission Rate	R_b	2.5 *Gbps*
CW Laser Power	P_T	10 *dBm*
Frequency range	f	193.5 to 192 THz
Channel spacing	___	100 GHz
Optical amplifiers gain at the transmitter	G_T	10 dB
Optical amplifiers gain at the receiver	G_R	10 dB
Transmitter's Apertures	D_T	5 *cm*
Receiver's Aperture	D_R	20 *cm*
Other losses	L_o	0 dB
Beam Divergence	θ_{BD}	2 *mrad*
Photodetector Type	___	PIN
Modulation Type	___	NRZ

Atmospheric Attenuation in Different Weather Conditions

The effect of different weather conditions on the performance of multiple transmitters/receivers in WDM-FSO links is considered at different visibilities for 1550 *nm*. It is found that the atmospheric attenuations of the FSO channel using the standard "Kim model" in fog, haze and clear weather conditions is as shown in Table 2

Table 2: The visibility and the atmospheric attenuation of Kim model.

Weather Condition		Visibility (*Km*)	Attenuation (*dB/Km*)
Fog	Dense	0.1	169.8
	Low	0.9	12.466
Haze	Dense	2	4.285
	Low	6	0.736
Clear	Clear Air	10	0.4416
	Very Clear	50	0.0883

And the Charbonneau's model is used for rainy weather condition is as shown in Table 3.

Table 3: The precipitation rate and attenuation values for rainy weather condition.

Weather Condition		Precipitation Rate (*mm/hr*)	Attenuation (*dB/Km*)
Rain	Light	2.5	1.982
	Strong	25	9.2

V. RESULTS AND DISCUSSIONS

This section, show performance of WDM-FSO communication systems with multiple transmitters/receivers under different weather conditions using the MIMO principle using OptiSystem version 7.0.

1) Data rate

In order to explore the effect of varying the data rate on the performance of WDM-FSO system with multiple transceivers, Figure 10 show the Q-factor of 16×1 WDM-FSO communication systems with multiple transceivers (1 beam, 2 beams and 4 beams) under dense foggy weather condition. In the figure, the Q-factor versus the transmitted power is compared with two values for the bit rates, 2.5 $Gbps$, 5 $Gbps$.

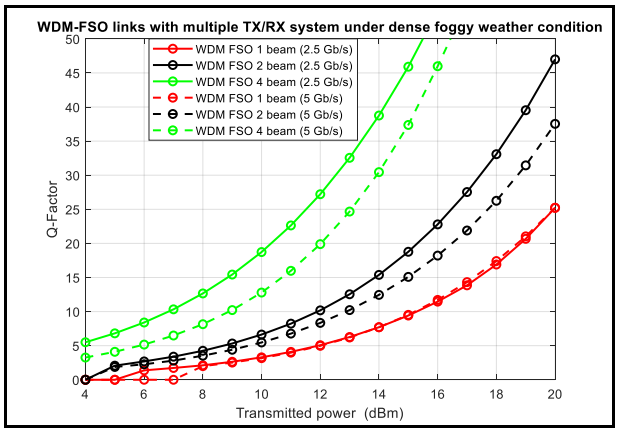


Figure 10: Q-factor versus the transmitted power of WDM-FSO links with multiple TX/RX system under dense foggy weather condition.

And the BER performance versus the transmitted power are compared with two values for the bit rates, 2.5 $Gbps$, 5 $Gbps$ shown in Figure 11.

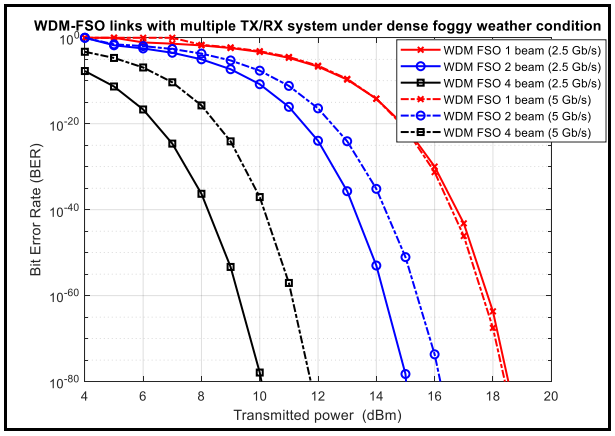


Figure 11: BER performance versus the transmitted power of WDM-FSO links with multiple TX/RX system under dense foggy weather condition.

It is observed from the previous figures under dense foggy weather condition, that increasing the number of transceivers improves the Q-factor and BER performance, and 16×1 WDM-FSO systems with lower bit rates for the same number of transceivers outperform the systems with higher bit rates. Moreover, the best results in both Q-factor and BER performance where the least bit rate with higher transverse is obtained from WDM FSO 4 beams (2.5 $Gbps$).

2) Transmitted power

In order to explore the effect of varying the transmitted power on the performance of WDM-FSO system with multiple transceivers, Figure 12 show the Q-factor of 16×1 WDM-FSO communication systems with multiple transceivers (1 beam, 2 beams and 4 beams) under strong rainy weather condition. In the figure, the Q-factor versus the link distance is compared with two values for the transmitted power,10dBm and15 dBm.

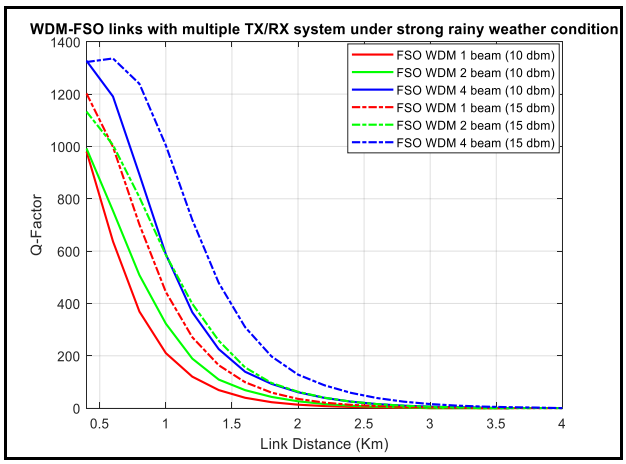


Figure 12: Q-factor versus the link distance of WDM-FSO links with multiple TX/RX system under strong rainy weather condition.

And the BER performance versus the link distance are compared with two values for the transmitted power, 10 dBm and 15 dBm shown in Figure 13.

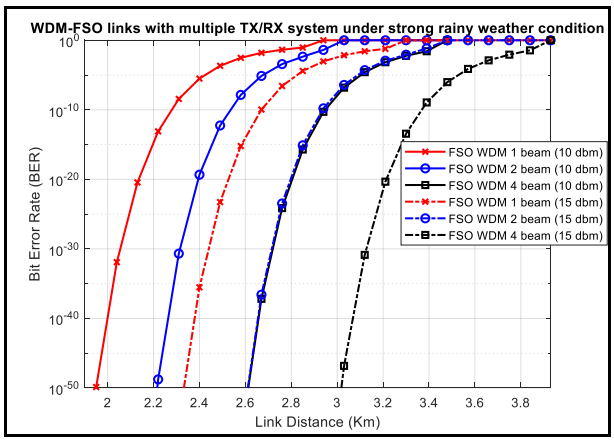


Figure 13: BER performance versus the link distance of WDM-FSO links with multiple TX/RX system under strong rainy weather condition.

It is clear from these figures for and strong rainy weather condition, that increasing the number of transceivers improves the BER performance and the Q-factor, and WDM-FSO systems with higher transmit power (15 dBm) outperform the systems with lower transmit power (10 dBm). Furthermore, the BER performance and Q-factor of WDM-FSO systems with higher transmit power and number of transceivers is better than that with lower values.

3) Receiver's Aperture

In order to explore the effect of varying the Receiver's Aperture on the performance of WDM-FSO system with multiple transceivers, Figure 14 show the Q-factor of 16×1 WDM-FSO communication systems with multiple transceivers of (1 beam, 2 beams and 4 beams) under dense foggy weather condition. In the figure, the Q-factor versus

the link distance is compared with two values for the Receiver's Aperture, 10 cm and 20 cm.

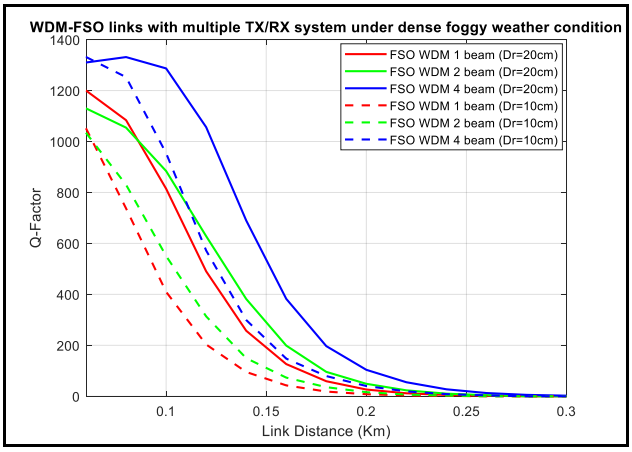


Figure 14: Q-factor versus the link distance of WDM-FSO links with multiple TX/RX system under dense foggy weather condition.

The BER performance versus the versus the link distance are compared with two values for the Receiver's Aperture, 10 cm and 20 cm. shown in Figure 15.

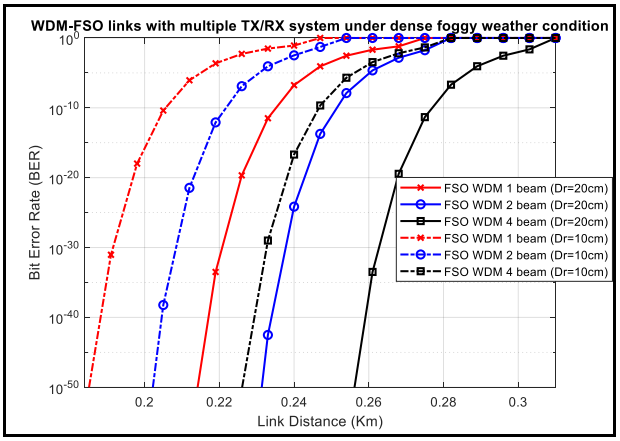


Figure 15: BER performance versus the link distance of WDM-FSO links with multiple TX/RX system under dense foggy weather condition

It is clear from these figures for dense foggy weather condition, that increasing the number of transceivers improves the BER performance and the Q-factor, and WDM-FSO systems with higher the Receiver Aperture (20 cm) outperform the systems with lower Receiver Aperture (10 cm). Moreover, the BER performance and Q-factor of WDM-FSO systems with higher Receiver Aperture and number of transceivers is better than that with lower values.

4) Maximum Link Distance in WDM-FSO MIMO Systems

It is clear from the previous results that the magnitude of the maximum Q-factor increases with increasing the number of transceivers in WDM-FSO MIMO systems. In addition, the BER performance and the Q-factor are improved as the value of the transmit power and Receiver Aperture increases.

The changes in the two values for the transmitted power, 10 dBm and 15 dBm. verify the improvement of WDM-FSO link performance with multiple transceivers under strong rainy weather conditions as shown in Tables 4 maximum link distance at nearly 10^{-9} BER.

Table 4: The maximum link distance of 16×1 WDM-FSO MIMO

Transmit power	maximum link distance (Km) at nearly 10^{-9} BER under strong rainy		
	1 × 1	2 × 2	4 × 4
10 dBm	2.31	2.51	2.98
15 dBm	2.69	2.96	3.4

The changes in the two values for the Receiver's Aperture, 10 cm and 20 cm. verify the improvement of WDM-FSO link performance with multiple transceivers under dense foggy weather conditions as shown in Table 5 maximum link distance at nearly 10^{-9} BER.

Table 5: The maximum link distance of 16×1 WDM-FSO MIMO

Receiver's Aperture	maximum link distance (Km) at nearly 10^{-9} BER under dense foggy		
	1 × 1	2 × 2	4 × 4
10 cm	0.21	0.224	0.249
20 cm	0.235	0.253	0.28

It is clear from previous Tables the maximum link distance of 4-TX/4-RX (4 beam) systems is greater than that in the case of 2-TX/2-RX (2 beam) and FSO systems with single transceivers (1 beam) due to the performance improvement achieved by increasing the number of transceivers.

VI. CONCLUSION

The performance of WDM-FSO communication links with Multiple Input Multiple Output (MIMO) directly depends on the distance used and also on atmospheric attenuation which is affected by different weather conditions such as fog, rain, and haze. The atmospheric attenuation of the WDM-FSO MIMO system is related to the visibility range, which is low in fog, moderate in haze and rain, and high in clear air weather condition.

The Q-factor and BER performance of WDM-FSO MIMO systems can be improved as the number of transceivers (beams) increases.

The performance of WDM-FSO MIMO communication system with lower bit rates (2.5 $Gbps$) and higher beams (4 beams) is better than system with higher bit rates (5 $Gbps$) and higher beams (4 beams). And this system under transmitted power of 15 dBm and (4 beams) showed best performance and highest maximum link distance compared to transmitted power of 10 dBm and other beams. The Receiver Aperture of 20 cm and (4 beams) showed best performance and highest maximum link distance compared to Receiver Aperture of 10 cm and other beams.

REFERENCES

[1] A. Prokes and V. Skorpil, "Estimation of free space optics systems availability based on meteorological visibility, "presented at IEEE Latin-American Conference on Communications, 2009.

[2] Malik, A., & Singh, P. (2015). Free Space Optics: Current Applications and Future Challenges. International Journal of Optics, 2015, 1–7.

[3] Lee S, Kwon JK, Jung SY, & Kwon YH (2012), Evaluation of visible Light communication channel delay profiles for automotive applications, Eurasip Journal Wireless Communication Network, Vol.12, no. 1, pp. 370-378.

[4] Khalighi, Mohammad Ali; Uysal, Murat (2014). Survey on Free Space Optical Communication: A Communication Theory Perspective. IEEE Communications Surveys & Tutorials, 16(4), 2231–2258.

[5] Alkholidi, A. G., & Altowij, K. S. (2014). Free Space Optical Communications Theory and Practices. Contemporary Issues in Wireless Communications.

[6] Nur Islam and Nur-AI-Safa Bhuiyan, "Performance Analysis and Investigation of Optimum Link Distance for Point-to-Point Free Space Optical Link in Different Weather Conditions," 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Savar, Bangladesh, May 21-23, 2015.

[7] Vanderka, A., Hajek, L., Bednarek, L., Latal, J., Vitasek, J., Hejduk, S., & Vasinek, V. (2016). Testing FSO WDM communication system in simulation software optiwave OptiSystem in different atmospheric environments. Laser Communication and Propagation through the Atmosphere and Oceans V.

[8] Al-Gailani, S. A., Mohd Salleh, M. F., Salem, A. A., Shaddad, R. Q., Sheikh, U. U., Algeelani, N. A., & Almohamad, T. A. (2021). A Survey of Free Space Optics (FSO) Communication Systems, Links, and Networks. IEEE Access, 9, 7353–7373.

[9] Nur Haedzerin Md Noor, Ahmed Wathik Naji, Wajdi Al-Khateed, "Performance Analysis of a Free Space Optics Link with Multiple Transmitters/Receivers," in IIUM Engineering Journal, vol. 13, pp. 49-58, 2012.

[10] Mahdi Hamzah, S, Ibrahim, A. Murdas, IA. Enhancement of the performance of DWDM free space optics (FSO) communications systems under different weather conditions. Int J Intell Eng Sys 2020;13.

[11] Singh, H., & Chechi, D. P. (2019). Performance Evaluation of Free Space Optical (FSO) Communication Link: Effects of Rain, Snow and Fog. 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN).

[12] Duvey D, Ritu Gupta Er. Review paper on performance analysis of a free space optical system. Int J Appl Innov Eng Manage (IJAIEM) 2014 Jun;3(6):135–9.

[13] Nazmul and S. Majumder "Performance Analysis of a free- space optical communication system through atmospheric turbulence channels," M. Sc thesis, December 2008.

[14] M. N. Islam and M. N. Al Safa Bhuiyan, "Effect of operating wavelengths and different weather conditions on performance of point-to-point free space optical link," International Journal of Computer Networks & Communications, vol. 8, no. 2, pp. 63–75, 2016.

[15] Shake, I., & Takara, H. (2002). Averaged Q-factor method using amplitude histogram evaluation for transparent monitoring of optical signal-to-noise ratio degradation in optical transmission system. Journal of Lightwave Technology, 20(8), 1367–1373.

Measurements of Electromagnetic Radiation from Mobile Base Stations Installed on Public Buildings at Tripoli Libya

Balam O. Huta
Department of electrical and electronic engineering
Tripoli University
Tripoli, Libya
b.huta@uot.edu.ly

Ibrahim M. Saleh
Department of electrical and electronic engineering department.
Tripoli University
Tripoli, Libya
ElhajAli@uot.edu.ly

***Abstract*— With the continuous use and proliferation of cellular mobile phones including other handheld communication devices to provide a robust support for radio signals coverage and good service quality provision to the connected loads (i.e. mobile users), has in turn steered the increasing deployment of Base Station "BS" antennas, denoting the antennas employed for installation on a BS or inside the handheld devices, therefore, it is essential to measure the radiation, to check the levels if they are within the permutable standard level, and measuring the radiation for the same zones to compare if the values have increased or decreased and calculate the difference over the years to estimate the relation between the increase of base station and its effect on the radiation (increase or decrease), in addition of, permitting future studies to compare with the obtained values. This research calculates the Specific Absorption Rate (SAR) from base stations in Tripoli, Libya and compare the values obtained with the limit values that world standards, such as ICNIRP and IEEE, suggests not to surpass for the possible biological harm.**

Keywords—radiation of power, compliance, ICNIRP, IEEE.

I. INTRODUCTION

Many studies were conducted in the last decades to observe the relation between exposure to EM radiations and health measured illnesses. Numerous publications stated that, health hazard is linked to the exposure of radiation. Many studies show that, radiation from base-stations is dangerous to health and in fact it agrees that, under some circumstances electromagnetic radiation may have determinately effect on human health. Many researches and studies about the effect of EM radiation towards the human organs, especially to the brain, which indicates the importance of the issue due to the presence of mobile phone when used in a close proximity to the head [4], [5] and [6].

This perturbing concern proffered Libyan researchers to conduct measurements to acquire mobile base station power values, namely, in 2007 a research conducted in Tripoli measured the power and electric field of mobile base stations installed in educational establishment [1], the researcher evaluated that the maximum values measured complied with the International Commission Non-Ionized Radiation Protection (ICNIRP)'s standard reference with an estimated of 94.32% compliance. The health concern detect to repeat the measurements frequently to ensure the compliance of EM radiation to the world level mainly ICNIRP's standard reference, regarding this concern it is essential to conduct the EM radiation measurements in the public premises to insure the compliance of measurements to the ICNIRP standard world level, and make a comparison with the values acquired in the previous research [1]. The measurements were taken in the years 2020 and 2021.

II. METHODOLOGY OF MEASUREMENTS AND SETUP

Measurements methodology in this research can be interpreted in the following:

A. Site Selection

These steps encompasses the partitioning of the province based on its temperament with regard to its convention that is, commercial division, industrial division, habitations division, educational division, and medical sectors [9], this steps is essential considering its aid in determining the sectors more impressionable to high radiation levels particularly, sectors patronized by elderly, and juvenile.

We targeted zones with: dense traffic base stations, medical institution with ill individuals, domestic sites where people spend the majority of their time in addition to educational establishments.

The studied sites are listed below:

- Domestic building.
- Central hospital, Tripoli, covid-19 isolation unit.
- Janzour high school.
- University of Tripoli, faculty of science, library.

B. Measurements

Measurements were performed multiple times in the course of the day as well as week and weekends. It is relevant to note the traffic disparity, for instance the uppermost traffic level measured was in the morning from about 8 am to 11 am, let alone traffic alteration throughout days of the week comprehensively, mobile base station traffic fluctuation is deliberated in the proceeding [7] and [8].

C. Acquired Data Analysis

The undertaken measurements were collected in multiple times of the day and collected during the days of the week as well as the weekends.

- Measurements were repeated two or three times on spot for the same point at the exact same moment for better accuracy.

- Peak time value is determined for each enterprise, noon, morning, or later in the day, this helps give a better comprehension of the process.

For each individual base station, we measured at:

1. Roof top: Under the tower, North, East, West, and south the tower at numerous distances.
2. Floors: If the building is composed of multiple floors, we estimated the power at each floor separately by measuring, under the tower, North, East, West, and south the tower at numerous distances.

- The measurements were repeated at least three times to ensure the precision and rightness of the values.
- The measurements of one were repeated on different days as well as on different times of the day.

D. *Instruments*

The SMP2 is an instrument for measuring electromagnetic fields. Its frequency range is determined by the field probe used. A range of frequencies compatible probes are available, from DC to 40 GHz [11].

III. MEASUREMENTS AND RESULTS

Measurements were taken at the end of 2020 and beginning of year 2021. The power was measured at sites with different circumstances namely, dense traffic found in Tripoli University (Library) and the number of operating base stations, the domestic nature of the area, and the health institution encompassing ill people.

The following illustrates the measured power at each site.

A. *Power density values measured at the domestic site (Maharat Company)*

Predominantly, measurements were conducted from 09:00 to 15:00 contingent on the working hours for three days in addition of the weekend in the period of 25-26-27-31/October/2020.

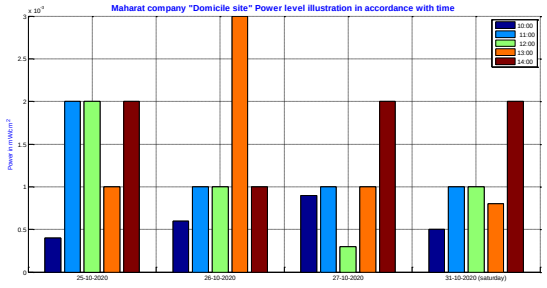


Fig. 1. A demonstration of the home site power level values measured at the roof of the building, attributable with the hours of the day with regards of the measuring days. The figure illustrates five different timings correspondent to the site.

Fig. 1. delineates the power variation in values accordingly with the hours of the day and in contradistinction to the measuring days, the maximum values are measured at 13:00 and 14:00 midday time. The maximum power measured (0.003 mW/cm^2) with an IEEE compliance equivalent of 0.11% and ICNIRP compliable of 0.75% [2] and [3].

B. *Power density values measured at Central hospital (Covid-19 isolation ward)*

The measurements were conducted in the interval of 07-23-25-26/November/2020 including a weekend day (Saturday), in the course of several hours of the day.

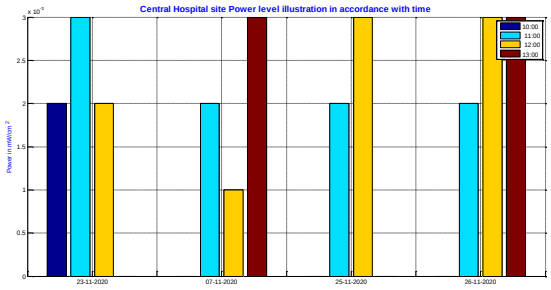


Fig. 2. A demonstration of central hospital (isolation ward) site power level values measured at the roof of the building, attributable with the hours of the day with regards of the measuring days..

Fig. 2. delineates the power variation in values accordingly with the hours of the day and in contradistinction to the measuring days. The maximum power measured (0.004 mW/cm^2) with an IEEE compliance equivalent of 0.14% and ICNIRP compliable of 0.94%.

C. *Power density values measured at the educational site (Janzour high school)*

Measurements were taken in the interval of 6,7,9/December 2020, the site includes two base stations installed at the roof of the school.

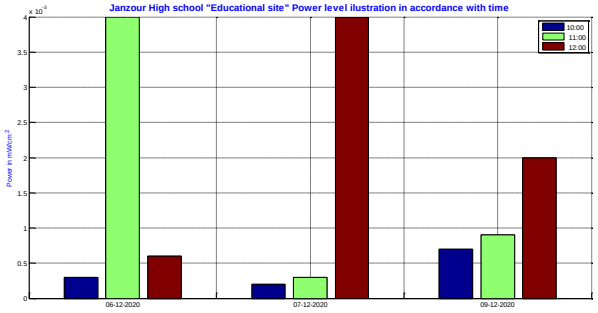


Fig. 3. A demonstration of educational institution (Janzour high school) site power level values measured at the roof of the building, attributable with the hours of the day with regards of the measuring days. The figure illustrates three different timings correspondent to the site.

Fig. 3. delineates the power variation in values accordingly with the hours of the day and in contradistinction to the measuring days. The maximum power measured (0.004 mW/cm2) with an IEEE compliance equivalent of 0.14% and ICNIRP compliable of 0.94%.

D. *University of Tripoli, faculty of science (library building)*

Measurements were taken on the rooftop of the building in the period of 05-06-07/January/2021. With a total of three base stations installed at different edges of the building.

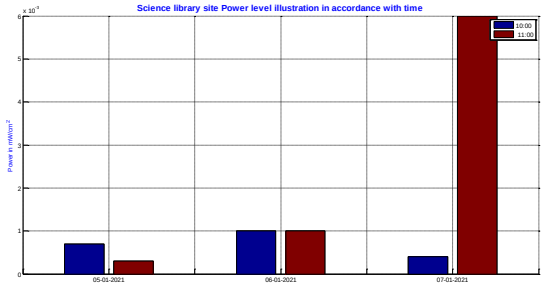


Fig. 4. An demonstration of Tripoli university faculty of science (library) site power level values attributable with the hours of the day with regards of the measuring days. The figure illustrates two different timings correspondent to the site.

Fig. 4. delineates the power variation in values accordingly with the hours of the day and in contradistinction to the measuring days, the maximum values are measured at 13:00 and 14:00 midday time. The maximum power measured (0.007 mW/cm^2) with an IEEE compliance equivalent of 0.27% and ICNIRP compliable of 1.75%.

E. Power levels measured at the rooftop with respect to each location consistent with the measuring days

Fig. 5. presents an illustration of the power measuring for each corresponding location accordant to the contingent measuring days, from the figure we can perceive the vicissitude in the power level in association with all locations, minimum values are perceived at the educational institutions (Janzour high school and the library) as opposed to the highest measured value 0.007 mW/cm^2 measured at the library site, for the same site.

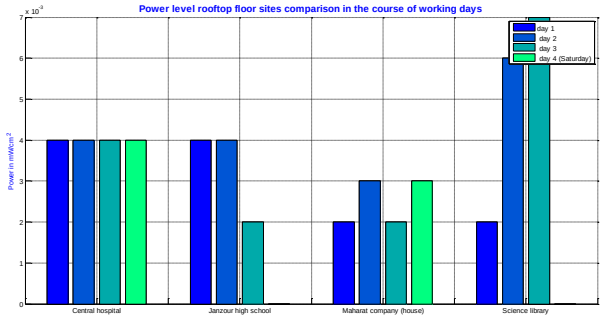


Fig. 5. Elucidate power levels measured at the rooftop corresponding each site in the course of the accordant measuring days.

F. Ground floor power values delineation with respect to all the selected sites in the course of the measuring days

Fig. 6 illustrates the power values measured at the ground floor corresponding each site in the course of the days of measurements, incontrovertibly the highest measured value was measured at the library site at Tripoli university, taking into account that the peak value measured appertaining to values measured at the rooftop amidst all values acquired was ascertained at the same site, the power decline from site to other compared to the peak value amounts to: the decline in the janzour high school is equivalent to50 %, this comes as no surprise that the second highest value is found on this site considering it encompasses two stations on its roof. Subsequently, the lowest measured value was measured at the domestic site, with a decline equivalent to 80%.

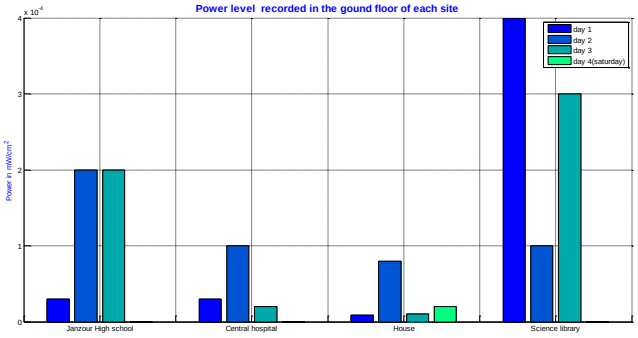


Fig. 6. A delineation of power levels measured at the ground level corresponding of each site in accordance of all measuring days.

G. Rooftop subsequent floor power values delineation with respect to all the selected sites in the course of the measuring days

Fig 7 demonstrates the power level measured at the subsequent floor of the rooftop in accordance with each site, abruptly, the peak value was measured at the domestic site the power decline from site to other compared to the peak value amounts to: the decline compared to the library is equivalent to 14.3%, this comes as no surprise that the second highest value is found on this site considering it encompasses three stations on its roof. Subsequently, the lowest measured value was measured at the high school site, with a decline equivalent to 85.7%.

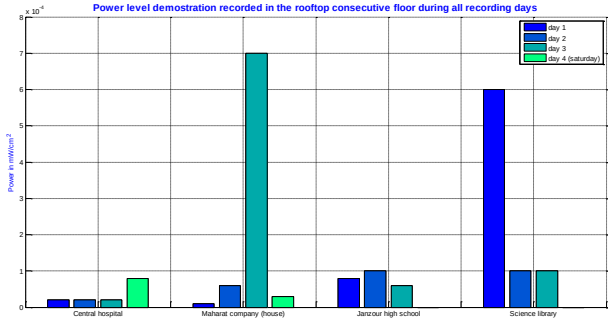


Fig. 7. A delineation of power levels measured at the rooftop subsequent floor corresponding of each site in accordance of all measuring days.

H. IEEE and ICNIRP SAR and power comparison

The Specific Absorption Rate (SAR) calculated from the measured power in accordance with the virtual family members and the 95th percent cuboid approach as referred to in [12], the calculated SAR to the maximum ICNIRP and IEEE permutable SAR. The highest percentage established in this research is 8.75%

Reference to [1], The maximum power obtained in 2007 was measured in Azzahf Al-Mostamer school with an equivalent of 22.546 µW/cm^2 representing 5.68% of the ICNIRP and IEEE limit value, comparing the maximum with obtained in this research which is 7 µW/cm^2 with a difference of 31.05%, on the other hand our values do not

deviate much, for instance, the lowest value measured is 3 μW/cm^2 with a difference in value of 57.14% whilst for the research the difference between maximum and minimum (22.546 and 0.576 μW/cm^2) values is 97.4%.

In order to make our research inclusive it is essential to compare our values to the values obtained in studies done in other countries. In the research conducted in Bulgaria in the year of 2018 [10] where the maximum power measured was (10 μW/cm^2 = 0.1 W/m^2) equivalent to 2.22% of the ICNIRP maximum values.

Table I. illustrates the maximum SAR calculated from the measured power in accordance with the virtual family members and the 95th percent cuboid approach, the percentage values given in the table represents the abidance of the calculated SAR to the maximum ICNIRP and IEEE permutable SAR. The highest percentage established in table 1.3 is 8.75% of the maximum SAR values allowed by IEEE the highest achieved in comparison with all other sites studied in this thesis, corresponding the whole-body SAR on account of the maximum SAR permitted is the lowest value equivalent to 0.08 W/kg.

In table II, it is observed that the higher the frequency the higher the power limit, for instance at 2100 MHz the measured power scored 0.08% and 0.39% of the maximum values of 10.5 W/m2. The maximum percentage achieved is 3.75% of the SAR value permitted by ICNRIP, the whole-body maximum SAR permitted is equivalent to 0.08 W/kg, the highest percentage established is 8.75% of the maximum SAR values allowed by IEEE.

TABLE I. COMPARISON BETWEEN ACQUIRED MEASURED SAR AND ITS COMPLIANCE TO IEEE AND ICNIRP REFERENCE VALUES.

Whole-body exposure *SAR (W/kg)*		**Calculated SAR (W/kg)**		***Comparison to general public SAR***	***Comparison to people in controlled environments***
General public	***People in controlled environments***				
0.08	0.4	VFM	0.00185	2.31%	0.4625%
		VFF	0.0023	2.87%	0.575%
		VFB	0.007	8.75%	1.75%
		95th percent cuboid	0.0021	2.62%	0.525%

TABLE II. MEASURED POWER IN COMPARISON WITH IEEE & ICNIRP ESTABLISHED MAXIMUM PERMUTABLE POWER VALUES FOR DIFFERENT RANGES OF FREQUENCIES

Operating frequency and the power correspondent to IEEE and ICNIRP limit (f_1 and f_2)[a]	***Percentage values to ICNIRP and IEEE f_M /200***
Operating frequency = 800 MHz, therefore f_1=20 and f_2= 4 W/m^2	1.75%
Operating frequency = 900 MHz, therefore f_1=22.5 and f_2= 4.5 W/m^2	1.56%
Operating frequency = 1800 MHz, therefore f_1=45 and f_2= 9 W/m^2	0.78%
Operating frequency = 2100 MHz, therefore f_1=52.5 and f_2= 10.5 W/m^2	0.67%
Operating frequency = 850 MHz, therefore f_1=21.25 and f_2= 4.25 W/m^2	1.65%

[a] $f_1=f_M/40$, $f_2=f_M/200$, where f_M is the frequency in megahertz.

IV. CONCLUSION

We measured the power and calculated the SAR for the mentioned locations: educational, health, and domestic sites in the period of: (end of year 2020 and the beginning of 2021). From the inclusive examination of the data, we were able to observe the following:

The measured power values are within the permutable range on account of the frequency range established in this research., in a detailed illustration, the highest compliance rate measured is equivalent to 99.9%, illustrating that the highest power measured is 0.1% of the values established by IEEE, In comparison with the least value equivalent to 1.75% of the established of IEEE power level. This reveals that all power measured at educational, health institution and domestic sites are compiled to the standards, therefor there are no danger to the people that is, excluding the long-term impact of exposure.

Comparing the values obtained in 2007 with the values obtained in this research we can state that: the maximum power obtained in 2007 was equivalent to 22.546 μW/cm^2 representing 5.68% of the ICNIRP and IEEE limit value, comparing the maximum with obtained in this research which is 7 μW/cm^2 with a difference of 31.05%, on the other hand our values do not deviate much, for instance, the lowest value measured is 3 μW/cm2 with a difference in value of 57.14% whilst for the research the difference between maximum and minimum (22.546 and 0.576 μW/cm2) values is 97.4%. This is because the mobile devices used today may have a lower power consumption in comparison to the mobile used in the time when the study in 2007 was done, and the instrument used to measure the power in this research is more accurate in comparison to the instrument used in the research done in 2007.

The comparison made with the research conducted in Bulgaria indicates that, our research gave a 1.75% of the ICNIRP permutable value at 900 MHz, the research in [12] gave a 2.22% of the ICNIRP permutable value, which indicates that power level are within the limit of exposure and are safe.

ACKNOWLEDGMENT

We, the authors, would like to thank Almadar company for kindly having procured the instruments and its accessories that was used for measuring the power in this research, and for supporting the research.

REFERENCES

[1] Fatima A. Ashkal, Ibrahim M. Saleh, “measurements of electromagnetic radiation from mobile phone base stations in Tripoli”, the Libyan international conference on electrical and electronic engineering LAICEEE 23-26/10/2010.

[2] "ICNIRP Guidelines for limiting exposure to time-varying electric, magnetic and electromagnetic fields", international commission on non-ionized radiation, health physics 74(4);494-522, 1998.

[3] " IEEE standard for safety levels with respect to human exposure to radio frequency electromagnetic fields, 3 kHz to 300 GHz, IEEE std C95.1", IEEE, New York, USA, April 2006.

[4] Wainwright P, "Thermal effects of radiation from cellular telephones", Phys. Med. Biol, vol. 45, pp. 2363-2372, 2000.

[5] Bernardi P. Cavagnaro M. Pisa S. Piuzzi E, "Specific absorption rate and temperature increases in the head of a cellular-phone user", IEEE Trans. Microwave Theory Tech, vol. 48, pp. 1118-1126, 2000.

[6] Hirata A., "tempertature increae in the human eyes due to near-field and far-field exposure at 900 MHz, 1.5 GHz, and 1.9 GHz", IEEE, vol. 47, pp. 68-76, 2004.

[7] Byung Chan Kim, Hyung-Do Choi, and Seong-Ook Park, "Methods of evaluating human exposure to electromagnetic fields radiated from operating base stations in Korea", 2008.

[8] Byung Chan Kim, "Two different techniques for evaluating electromagnetic fields radiatioed from radio base stations in Korea", IEICE, 2009.

[9] Cesar Camilo Rodrigues, Celso Andres Forero, and Homero Otega Boada, "Electromagnetic field measurements method to generate radiation map", IEEE, 2012.

[10] Iv. Topalova1, Ts. Shalamanova1, V. Zaryabova1 and M. Israel1, “Electromagnetic field exposure from telecomunication sources in areas with “sensitive buildings and places”, international conference on radiation and application in variouse fields of research, vol. 3, pp 197-201, Mecedonia, 2018.

[11] https://www.wavecontrol.com/rfsafety/en/products/emf-meter-smp2.

[12] Marie-Christine Gosselin et all, "Estimation formulas for the specific absorption rate in humans exposed to base-station antennas", IEEE transactions on electromagnetic compatibility, 2011.

Effect of dust and sand storm on 5G Comunications in Libya

Ibrahim M.. Saleh
Department of Electrical and Electronic Engineering
University of Tripoli
Tripoli, Libya
ibrahimsaleh1950@gmail.com

Ahmed. S. Alatrash
Department of Electrical and Electronic Engineering
University of Tripoli
Tripoli, Libya
Ah.Alatrash@uot.edu.ly

Abstract— 5G is the next generation for communications and data transmit ion, as it is much faster , more capacity, and better network efficiency. Capacity could reach 100 times as that of 4G, propagation of millimeter waves could be severely affected by claimant conditions ,like rain , dust and sand particles , in terms of power degradation and wave polarization . unfortunately 5G transmission could only last few hundred meters ,and when pass through a media of dust and sand could be attenuated very much, and could lead to a shorter distance signal covers.
This paper is to study the effect of dust and sand storms on 5G signals when propagating through a media of sand and dust particles, in some suburb areas of Libya at 28GHz and 60 GHz. The study will also include effect of visibility and humidity on 5G signals in a region of sand and dust storms. a mathematical model will be used to predict dust and sand storms attenuation in the case of worst visibility of 4 meters, and a particle diameter of 50 micrometers. The results showed that a sever attenuation take place when propagating 5G signals- in wind storms.

Keywords— Millimeter-waves, Sandstorm, Attenuation, Propagation-wave, Transmission Loss, Humidity

I. INTRODUCTION

Dust and sand storms are a common climatic phenomenon in some suburb areas in most of north Africa [1]. The storm caused by strong winds will cause carrying sand and dust particles in the air media and this will decrease visibility due to high condensed particles. The contents of the dust and sand changes from a place to another, the chemical analysis of such particles could be iron, aluminum, silicon and or other alloys, these particles will change the media permittivity to a complex permativity , with its real part and imaginary part of the dielectric constant, in turn this will cause signal attenuation. The attenuation is function of frequency and increases for low visibility and more in case of humidity.

To be able to calculate the attenuation of 5G signals due to climate condition such as, dust and sand storm, and humidity. In a previous study in suburb areas [1] , we selected from a suburb area in Libya called Dlaim, samples from the hanging dust and sand in the area of study were collected, then chemical analysis of the samples was carried out, from which the average particle sizes was determined using some spatial instrumentations. In this study we used the dielectric constants provided from the previous study.

II. RESEARCH METHODOLOGY

This study is carried out on suburb areas in Libya for frequencies 28 and 60 GHz, to estimate the impact of dust and sand storms of 5G communication signals by predicting and analyzing the dust and storm effect, first we defined Tx and Rx parameters such as gain, frequency, transmit power and receiver sensitivity shown in table 1 , secondly we assumed that the channel propagation is to be under the worst-case scenario including free space, where the power received is computed under the influence of dust and sand storm. Finally, compared the received power with the receiver sensitivity which we assumed to be -85 dBm the received power computed based on the following parameters: transmitter power Pt, the gain of the transmitter and receiver antenna, Gt and Gr respectively, dust and sand storm attenuation L_D, free space loss L_{FS} , as shown in Table :1 .
Thus, we can calculate the received power by [2]:

$$P_R(dBm) = P_T + G_T + G_R - L_{FS} - L_D \quad (1)$$

Where the free space L_{FS} can calculate by:

$$L_{FS}(dB) = 32.44 + 20\log(f\ MHz) + 20\log(D\ Km) \quad (2)$$

Where f is the frequency of the 5G signal in (MHz) and D is the distance between the transmitter and receiver in Km

Table 1. System specification and parameters.

Description		Scenario 1	Scenario 2
Operating frequency		28 GHz	60 GHz
Transmitter	Antenna Power	10 dBm	15 dBm
	Antenna Gain	25 dBi	25 dBi
Receiver	Antenna sensitivity	-80 dBm	-80 dBm
	Antenna Gain	25 dBi	25 dBi

Now, to compute attenuation due to dust storm generally by solving the forward scattering amplitude function of a single particle. The solution may be found by using the Rayleigh approximation or Mie solutions. The method depends largely on the particle number and particle radius, the entire attenuation loss, in dB, caused by a dust storm over a link has a length of d can be given as [3]:

$$L_D = \int_0^d A_d dx \quad (3)$$

where A_d (dB/km) is the specific attenuation characterizing the dust and sand storm which can be expressed as:

$$A_d \left(\frac{dB}{Km}\right) = 4.343 \times 10^3 \int_{a_{min}}^{a_{max}} \sigma_t(a) \, . N(a) \, da \quad (4)$$

Where N(a)da is the number of particles per unit volume of air with dust particles radius between a and a + da, σ_t is the total attenuation cross section efficiency factors of dust particle of radius a, and to calculate the attenuation by the above equations, data are required for the number of particles of dust N, which is difficult to measure accurately, we can express the particle density in term of the visibility as [4]:

$$N = \frac{5.5 \times 10^{-4}}{V \, a_e^2} \quad (5)$$

After solving equation of attenuation, visibility and particle radius [3] we can express the specific attenuation due to dust and sand storm A_d (dB/Km):

$$A_d = \frac{a_e f}{V} \, (X + Y a_e^2 f^2 + Z a_e^3 f^3) \quad (6)$$

Where:
a_e: the equivalent particle radius in meters,
f: the frequency in GHz and
V: the visibility in kilometer,
X, Y and Z: constants whose values depend on real (ε') and imaginary (ε'') of the dielectric constant:

$$X = \frac{1886 \, \varepsilon''}{(\varepsilon'+2)^2 + \varepsilon''^2} \quad (7)$$

$$Y = 137 \times 10^3 \, \varepsilon'' \left(\frac{6}{5} \frac{7\varepsilon'^2 + 4\varepsilon' - 20}{[(\varepsilon'+2)^2 + \varepsilon''^2]^2} + \frac{1}{15} + \frac{5}{3[(2\varepsilon'+3)^2 + 4\varepsilon''^2]} \right) \quad (8)$$

$$Z = 379 \times 10^4 \left(\frac{(\varepsilon'-1)^2(\varepsilon'+2) + [2(\varepsilon'-1)(\varepsilon'+2) - 9] + \varepsilon''^4}{[(\varepsilon'+2)^2 + \varepsilon''^2]^2} \right) \quad (9)$$

We get the Complex Permittivity for nine samples in this study region from table 1 [1]:

Table 2. The complex permittivity and PSD for 9 sample

Sample No.	PSD	complex permittivity
1	sand	5.0384 - j 0.0509
2	sand	5.4851 - j 0.0562
3	dust	5.4801 - j 0.0694
4	sand	7.5929 - j 0.1140
5	dust	6.7899 - j 0.1296
6	dust	5.4003 - j 0.0787
7	dust	7.4707 - j 0.1344
8	dust	5.5713 - j 0.0704
9	dust	8.3078 - j 0.1329

The average complex permittivity of all samples is equal to 6.3485 - j 0.0929.
When the media includes moisture contents, the complex permittivity can be estimated from the following equations, with the variation of complex permittivity and relative humidity [5].

$$\varepsilon'_H = \varepsilon' + 0.004H - 7.78 \times 10^{-4} H^2 + 5.56 \times 10^{-6} H^3$$

$$\varepsilon'_H = \varepsilon'' + 0.002H - 3.71 \times 10^{-4} H^2 + 2.76 \times 10^{-6} H^3$$

III. Result and DiscussionS

To find the impact of the dust and sand storm on 5G signals in suburb areas of Libya, we obtained data from meteorological stations [2], so we can calculate the attenuation with variation of the visibility for 4 sample from table2, Fig1 and 2 show the attenuation due to visibility at 28 and 60 GHz respectively , noted that when the visibility increase the attenuation will decrease, tables from 3 to 8 describe the effect of humidity for the 4 sample at 60% , and 100% humidity ,as we see from table the attenuation increases with the humidity and decrease with visibility.

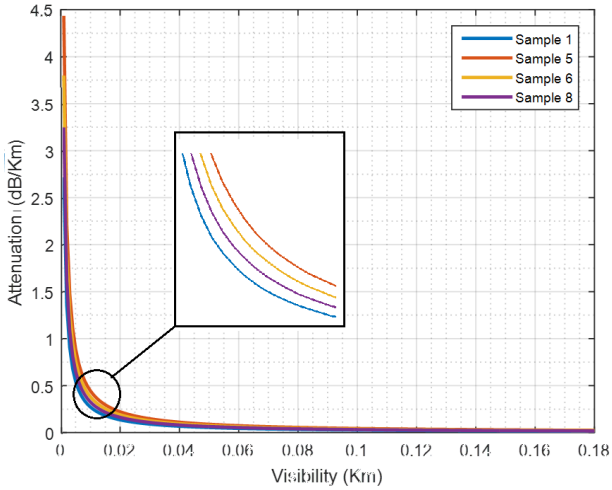


Fig 1. The power received versus distance between Tx and Rx at 28 GHz in free space loss.

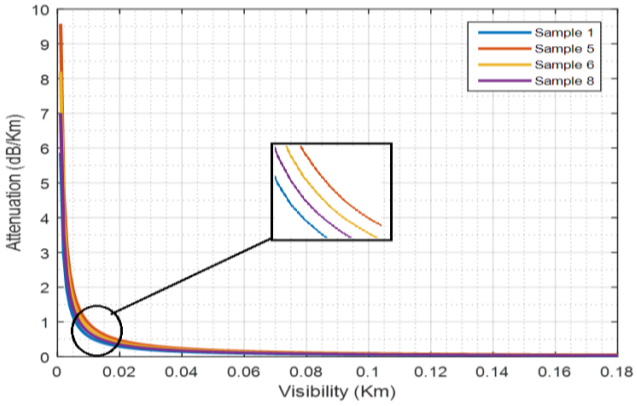


Fig 2. The power received versus distance between Tx and Rx at at 60 GHz in free space loss.

Table 3. Variation of visibility with sand-storm attenuation (dB/km) at 28 GHz, with humidity 0 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	0.679	0.181	0.027
5	1.109	0.296	0.045
6	0.949	0.253	0.038
8	0.827	0.221	0.033

Table 4. Variation of visibility with sand-storm attenuation (dB/km) at 28 GHz, with humidity 60 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	0.832	0.222	0.033
5	1.201	0.321	0.048
6	1.086	0.290	0.043
8	0.897	0.239	0.036

Table 5. Variation of visibility with sand-storm attenuation (dB/km) at 28 GHz, with humidity 100 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	0.931	0.248	0.037
5	1.265	0.338	0.051
6	1.174	0.313	0.047
8	0.950	0.251	0.038

Table 6. Variation of visibility with sand-storm attenuation (dB/km) at 60 GHz, with humidity 0 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	1.466	0.391	0.059
5	2.394	0.638	0.096
6	2.049	0.546	0.082
8	1.789	0.477	0.072

Table 7. Variation of visibility with sand-storm attenuation (dB/km) at 60 GHz, with humidity 60 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	1.795	0.479	0.072
5	2.598	0.693	0.104
6	2.342	0.625	0.094
8	1.939	0.517	0.078

Table 8. Variation of visibility with sand-storm attenuation (dB/km) at 60 GHz, with humidity 100 % for 4 sample.

Sample No.	Signal Attenuation (dB/km) with visibility		
	4m	15m	100m
1	2.007	0.535	0.080
5	2.731	0.728	0.109
6	2.532	0.675	0.101
8	2.037	0.543	0.081

The received power for the average complex permittivity of all samples as shown in table2 can be found, with the effect of dust\ sand storm, and free space attenuation, we choose the worst case for this study the visibility is considers to be 4m as given from [2], and particle diameter 50 μm .

Figure 3 shows the received power in the wireless channel at 28 GHz and 60 GHz respectively under the influence of free space attenuation versus the distance between the transmitter and the receiver, with sensitivity of the receiver is -80 dB, so that the signal can be received at distance 8.5 km and 7 km for 28 and 60GHz respectively.

Fig 4 shows the received power in the wireless channel at 28 GHz under the influence of free space attenuation with and without the dust storm attenuation against the distance between transmitter and receiver, as we mentioned before that the sensitivity of the receiver is -80dBm so that the signal can be received at 8.5 km without dust storm attenuation and with dust storm attenuation of the signal can be received at 5 km.

The received power in the wireless channel at 60 GHz will have attenuation effect more than that for 28 GHz , Figure 5 Shows the effect of free space attenuation with / without dust storm attenuation versus the distance between the transmitter and the receiver, with sensitivity of the receiver is -80 dB , the signal can be received at 7 km If there is no dust storm attenuation and with dust storm attenuation of the signal can be received at 3.4 km.

Figure 6 shows the received power in the wireless channel at 28 GHz under the influence of the free space attenuation with the attenuation of the dust storm and the effect of humidity coefficient at 0, 60, and 100 percent respectively against the distance between the transmitter and the receiver in this case the signal can be receiver at distance 5.06, 4.87, 4.75 Km .

Figure 7 shows the received power in the wireless channel at 60 GHz under the influence of the free space attenuation with the attenuation of the dust storm and with the effects of the humidity coefficient of 0, 60, and 100 percent respectively against the distance between the transmitter and the receiver, and when the receiver sensitivity is -80 dB the signal can be received at 3.1 km, 3.2 km, and 3.4 km respectively in all cases of humidity

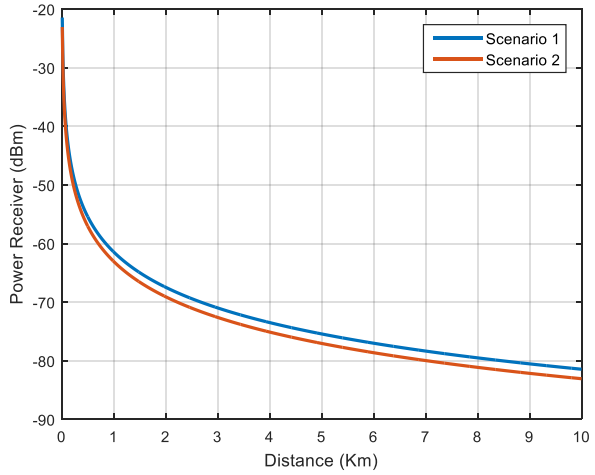


Fig 3. The power received versus distance between Tx and Rx at scenario 1 [28 GHz] and scenario 2 [60 GHz] with free space loss.

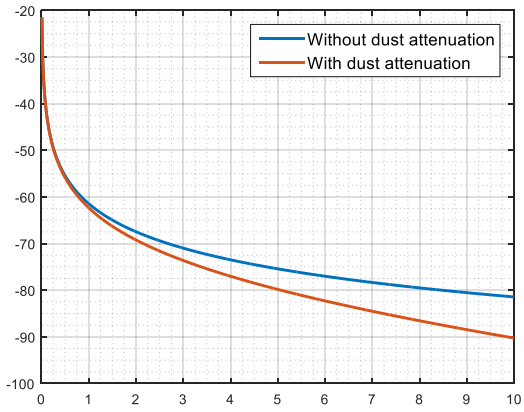


Fig 4. The power received versus distance between Tx and Rx at 28GHz with free space and dust storm attenuation.

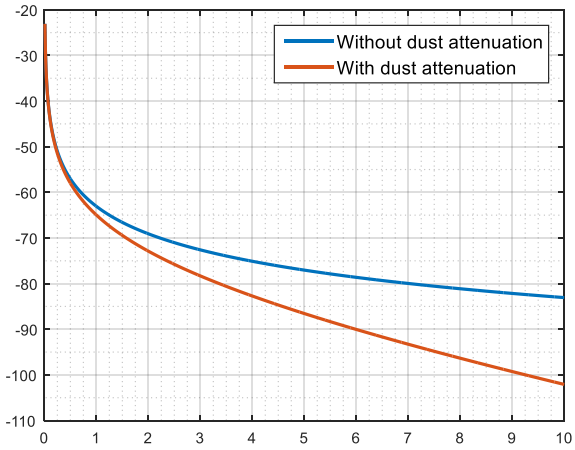


Fig 5. The power received versus distance between Tx and Rx at 60GHz with free space and dust storm attenuation

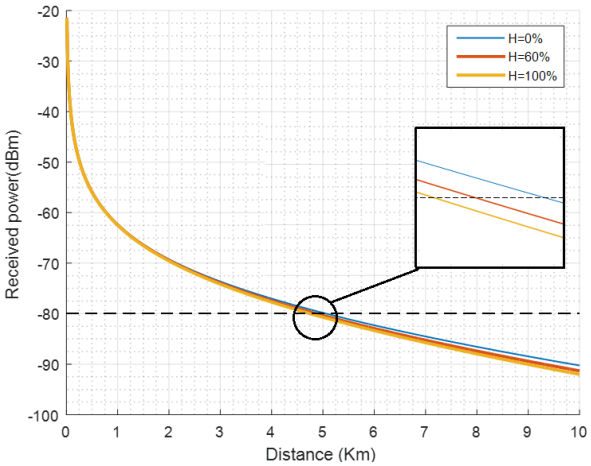


Fig 6. The power received versus distance between Tx and Rx at 28GHz with effect of humidity

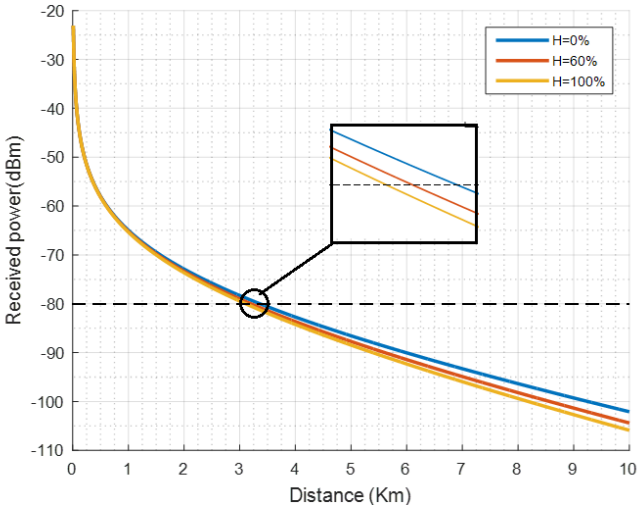


Fig 7. The power received versus distance between Tx and Rx at 60GHz with effect of humidity

From all above figurs it can be clearly seen that the the sand and dust has a great infulwance on 5G signales and it gets worst with humidity.

IV. Conclusion

5G transmission and mobile coverge in suburb areas of Libya was investigated for the effect of dust and sand storms at 28 and 60 Ghz , the study was based on prevous data measured for dust and sand storm done for a village named Dlaim . In this study we used a mathematical model to predict the attenuation introduced by dust and sand storm, the results showed that the signal is severly attenuated due to dust sand storm, and is higer for 60 GHz, low visibility increase attenuation, and so higher humidity. The coverge distance for 5G signal with dust and sand storm may drop to half its value in free space for rhe worst case of visibility and humidaty for 60 GHz, and to 80 % for 28 GHz.

References

[1] E. M. Abuhdima and I. M. Saleh, "Effect of sand and dust storms on GSM coverage signal in southern Libya," 2010 International Conference on Electronic Devices, Systems and Applications, Kuala Lumpur, 2010, pp. 264-268, doi: 10.1109/ICEDSA.2010.5503063.

[2] Propagation Data and Prediction Methods Required for the Design ofTerrestrial Line-of-Sight Systems, document ITU-R P.530-17, 2017.

[3] Z.E.O., Elshaikh, M.R., Islam, O.O.Khalifa, and H.E., Abd-El-Raouf, 2009. "Mathematical model for the prediction of microwave signal attenuation due to duststorm". Progress In Electromagnetics Research, 6, pp.139-153.

[4] Goldhirsh, J., "Attenuation and backscatter from a derived two-dimensional duststorm model," IEEE Trans. Antennas Propagation, Vol. 49, No. 12, 1703–1711, 2001.

[5] S.M., Sharif, "Attenuation properties of dusty media using Mie scattering solution." 2015. Progress In Electromagnetics Research, 43, pp.9-18.

Internet Usage Patterns and Traffic Analysis in Libya using Deep Packet Inspection Tools

Abdulmoaz Layas
Technical Department
Libyan International Telecom Co.
Tripoli, Libya
abdulmoaz.layas@litc.ly

Adam Asbali
Technical Department
Libyan International Telecom Co.
Tripoli, Libya
adam.asbali@litc.ly

Tammam Benmusa
Electrical and Electronic Eng. Dept.
University Of Tripoli
Tripoli, Libya
t.benmusa@uot.edu.ly

***Abstract*— Today, one cannot think of life without the Internet. The demand on the Internet has grown at a very fast pace, which has resulted in heavy Internet traffic. The Average traffic load has risen, and data traffic patterns have also become unpredictable. Therefore, network traffic monitoring and analysis for the Internet Service Providers (ISP's) have become essential to understand the user's usage behavior, troubleshoot and resolve problems effectively when they occur, then provide new commercial offers based on analysis outputs. All the Internet traffic from/to Libya is going through the Libyan International Telecom Company (LITC) network, which is representing the only gateway to the world through different ports. In this paper, the Libyan Internet traffic data was gathered during 2021-2022 for long period of time using a Deep Packet Inspection (DPI) tool installed on the LITC network. These data were analyzed and processed to obtain models for daily and weekly internet usage pattern for Libya. The patterns were obtained in three different periods of time to look at the effect pattern changing over the seasons. The results were compared with other patterns obtained from different works in other places.**

Keywords—Deep packet inspection, LITC, Libyan international telecom company, DPUP, WPUP, Internet usage pattern, Traffic analysis, ISP.

I. INTRODUCTION

Network analysis is a process of capturing network traffic and inspecting it closely to determine what is happening in the network. It also known by several other names such as network analysis, protocol analysis, packet sniffing and packet analysis. We consider network traffic analysis to be a set of methods that successively is utilized to understand the nature of traffic per packet basis [1].

Analyze traffic in a communication network is needed for many types of network research studies. If the target is to improve the network, or to be aware of the operational aspects of the network, the main purpose of traffic analysis is sometimes to find out the usage on different parts of the network. It includes monitoring user behavior or Internet Service Providers (ISP's) and observing how the networks operate under the traffic loads at different times [2].

There are several methods to acquire packets for network traffic analysis. The most used two methods are, NetFlow and In-line solution. NetFlow is taking samples from the whole traffic passing through the gateways such as Solarwinds System [3]. The second method is In-line solution method, which is physically inserting a network tap, duplicating, and sending the data stream to the analyzer tool for inspection. The inspection is done by certain technique, such as Deep Packet Inspection (DPI) Systems.[4]

DPI is a real-time network filtering and Internet traffic analyzing technology that mainly works in High-Speed network connection [5]. DPI can be implemented in the application layer of Open System Interconnection (OSI) model. It is called deep inspection because the inspection not only includes the packet headers but also covers the packet payloads [6]. "DPI technologies are intended to allow network operators precisely to identify the origin and content of each packet of data that passes through the networking hubs" [7].

In this paper the traffic data is gathered from the Libyan International Telecom Co. (LITC) network, which is the main and only internet gateway in Libya.

The gathered data has been extracted and collected using Huawei DPI tool during different months in daily basis. The data is collected for consecutive days for different months. The hourly averages were calculated per day, and the daily usage pattern in Libya was drawn. In addition, weekly data for different seasons over a year was collected to obtain the weekly usage pattern. All the extracted data and obtained patterns have been investigated and analyzed to insight into network and user's behavior, and to be used to observe the utilizations for the company links. Finally, the obtained results are compared with other available previous works for other countries.

This work is sponsored and approved and by (LITC). It may use outcome and the results of this paper to enhance its network and suggest new commercial offers for their clients.

II. THE LIBYAN INTERNATIONAL TELECOM CO. NETWORK

The use of internet is a main need of all people nowadays, almost you can do everything online, such as streaming videos, online bank services or share files. LITC is providing the wholesale IP/data internet service, wholesale international voice calls and cloud services to all other organizations and ISP's in Libya. Customers in Libya subscribe with Internet Service Providers (ISP's) such as Aljeel Aljaded, and Libya Telecom and Technology (LTT) who get their access the Internet through LITC.

The LITC international connections, shown in Fig1. It has three submarine routes connecting Libya internationally, and each route consist of several submarine fibers pairs. In Tripoli, there are two routes. First one Tripoli-Mazara which is connecting Libya to Italy and used as the main link of internet traffic. The second one connecting Tripoli to the Europe India Gateway (EIG) submarine cable, which is connecting Tripoli to London, Monaco, and Egypt, and used as a backup link for the internet traffic.

Fig 1. LITC international connections

Moreover, LITC has (Silphium) submarine cable, which is landed in Derna city and connecting Libya to Greece. Most of the internet traffic of LITC network is carried over the main route (Tripoli-Mazara), which contains two 100G

links in two different wavelengths. The traffic is divided equally between them with approximately 70% utilization of the total capacity for each link.

The features and the information about the international Internet traffic is monitored and inspected by Deep Packet Inspection (DPI) tool installed in number of submarine links.

III. Methodology

The traffic of Tripoli-Mazara route is chosen for this work since it is representing the main international route. The gathered data is taken from one of the 100G link in this rout, since the traffic of the two links in this route are almost identical.

The measurements were taken every short intervals and averaged every hour. The measurement system also gives the peak traffic value for every hour. It has been noticed that the shape of the daily pattern for the peek values and the average values are the same. The only different was the numbers in the graph. In this work, it has been focused on the hourly peak values, since they are the most important measure in analyzing the utilization issues.

The data under study was taken for 30 consecutive days for three different periods of a year. And the same process has done for each period. The reason of taken these months is to look at the pattern at different times of the year.

During this work, the staring time of the day was taken 05:00 AM to 04:59 AM of the next day, that was taken to make the peaks of the patterns uniform and more clear.

The monitoring and processing work done for daily basis and weekly basis, as per the following steps:

A. Daily Basis Patterns.

- Thirty daily reports have been extracted from DPI tool for each period.
- The peak value of traffic was taken for each hour during a day, which represents the Daily Peak Usage Pattern (DPUP).
- From the data of the first period, the peak values of the first hour of each day were taken for the thirty days of the period, and the average of these values was considered as the peak value of the first hour of the day.
- The step above was repeated for the other 23 hours (the 2nd, 3rd, ...up to the 24th hour) to obtain the average peak values for all 24 hours of the day.
- From these 24 average peak values, the average of DPUP was drawn.
- This was repeated for each period under study.
- The obtained average DPUP was compared with other available previous works.

B. Weekly Basis Patterns.

- Four weekly reports have been extracted from DPI tool in each period.
- The peak values in one hour interval were taken and drawn for the four weeks (28 days), which represents the Weekly Peak Usage Pattern (WPUP)
- Since we have four weeks, we have four peak values for each hour in each day of the week. These peak values were averaged and drawn to represents the average of WPUP.
- This was done for the three periods under study.

C. Capacity Utlization.

The utilization of the total link capacity were calculated for each day, in order to give a clear picture to the system capacity, which is very useful for the decision maker in the LICT company. This was done by taking the average values of the peaks in every day of the week and divided by the total link capacity to get the percentage of the link utilization.

IV. LITC Internet Traffic Analysis and Results

A. Daily Peak Usage Pattern.

The DPI tool installed on one of the 100 G link of LITC between Tripoli Mazara has used to get the traffic for three periods, each of them consist of thirty consecutive days. These periods were on January, June, and September. The monitoring and analyzing process described in section III A were implemented. And the obtained average DPUP was presented in Fig 2. Fig 3, and Fig 4 for the three period under the test.

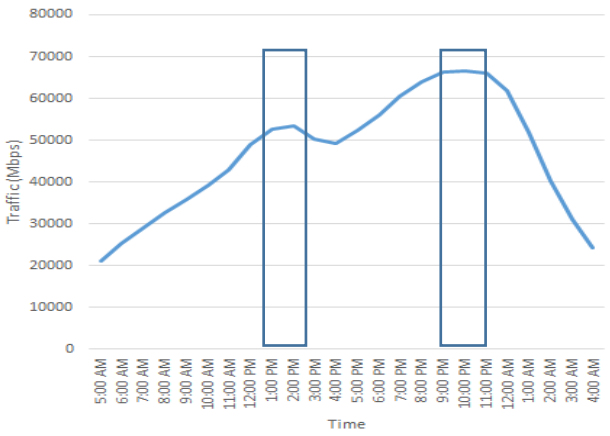


Fig 2. Average daily peak usage pattern in January.

The usage patterns of the three period are similar, since all of them have two peaks, first one occurs at the early afternoon where the second peak occurs at late evening and early morning. For winter season (January period) the first peak starts at 12.30 PM up to 2.30 PM, with a peak value of just above 50 Gbps. The second peak starts at 10.0 PM up to 1.0 AM, with a peak value around 66 Gbps.

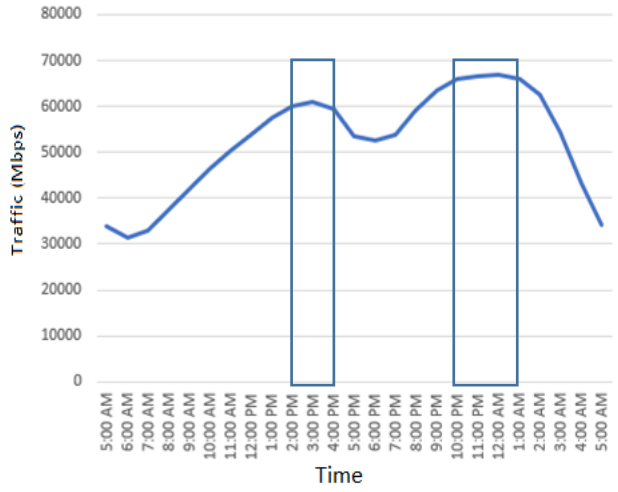


Fig 3. Average daily peak usage pattern in June.

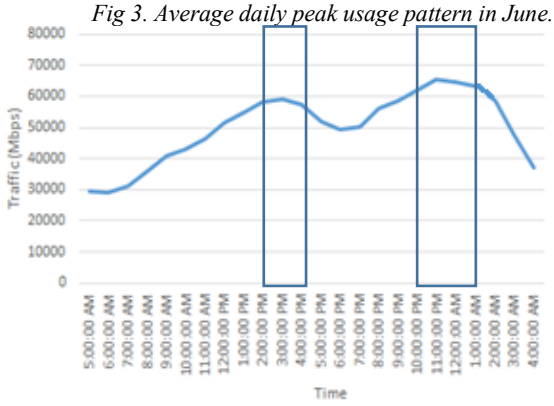


Fig 4. Average daily peak Usage Pattern in September

The usage patterns of the three period are similar, since all of them have two peaks, first one occurs at the early afternoon where the second peak occurs at late evening and early morning. For winter season (January period) the first peak starts at 12.30 PM up to 2.30 PM, with a peak value of just above 50 Gbps. The second peak starts at 10.0 PM up to 1.0 AM, with a peak value around 66 Gbps.

For summer season (June period) the first peak comes between 2.0 PM and 4.0 PM with peak value just above 60 Gbps, the second peak starts at 10.0 PM and stays until 1.0 AM, with a peak value around 66 Gbps, which is similar to that in wintertime with one hour longer. This is expected, since people in summer go to bed in later due to the hot weather.

The fall season (September time) pattern is very similar to summer pattern. The two peaks values, durations and occurring times are similar to those obtained from June measurements, as shown in Fig. 4.

In the three graphs, we can notice that the peak values dose not exceeds 70 Gbps. That is mean the maximum link utilization does not reach 70% of the total link capacity all over the year.

In order to compare the obtained results with other work results, the traffic values in Fall graph (Fig. 4) were redrawn in Fig. 5 by making the starting time of the day 12.00 AM. The pattern now was compared with different patterns obtained from another work measured in the same season (fall) in North America and UK [9] shown in Fig 6.

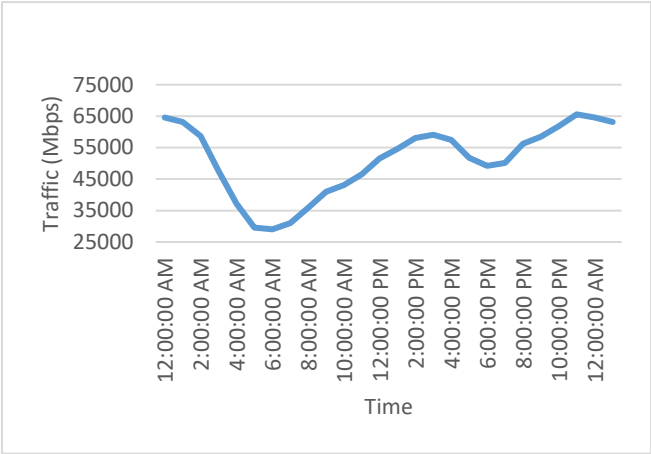


Fig 5. Daily peak usage pattern in Libya during Fall season

Comparing Fig 5 and Fig 6 we can notice the shape of the usage behaviors are almost the same in the three countries apart from the values of the peaks. Also, we can notice that the peaks in the other country graphs, spatially in UK graph, are combined and become one long peak.

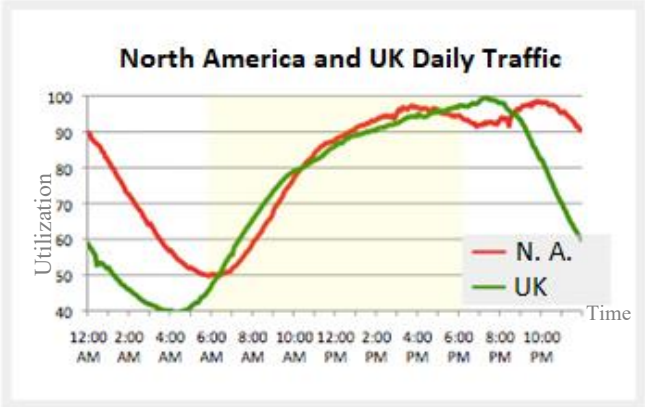


Fig 6. Daily internet traffic for North America and Europe [9].

B. Weekly Peak Usage Pattern..

The same DPI tool is used to collect the traffic data for this section from the same Tripoli Mazara 100 G link. Also the data collected for three periods as well, January, June, and September. The monitoring and analyzing process described in section III B were implemented. And the obtained average WPUP was presented in Fig 7. Fig 8, and Fig 9 given below. Please not the shifting of the starting time of the day 5.00 AM for the reason mentioned earlier in section III.

From above results we can see that the daily user behavior during seven consecutive days is similar in Winter and Fall, and in Summer we noticed that Friday has the least average peak traffic values, which is a holiday, and this result can be justified due to the Libyan social culture during Summer time, since most of the people go outside their homes where no internet access.

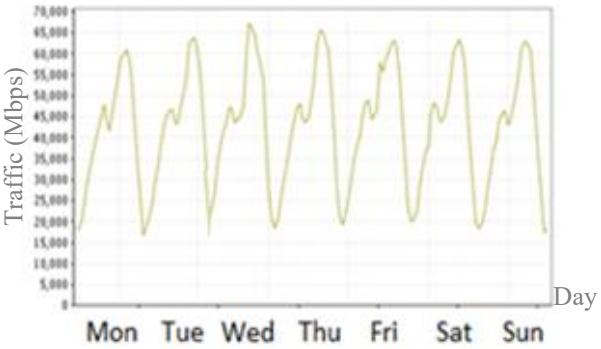


Fig 7. Average of weekly peak usage pattern in January

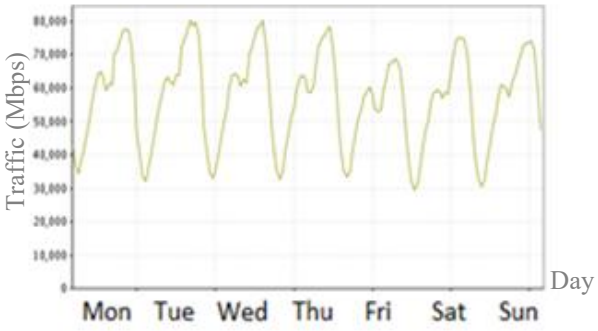


Fig 8. Average of weekly peak usage pattern in June

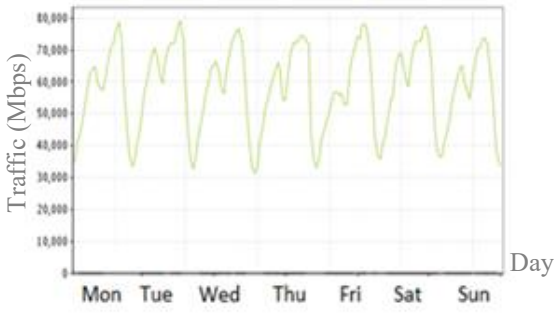


Fig 9. Average of weekly peak usage pattern in September

C. Capacity Utlization.

The link capacity utilization for the seven days of the week were calculated as described in section III C. The results were given in table 1 and drawn in Fig. 10.

Table 1. Table of days Vs traffic utilization

Day	***Utilization***
Mon	77.48 %
Tue	80.13 %
Wed	80.28 %
Thu	78.09 %
Fri	68.60 %
Sat	74.95 %
Sun	73.92 %

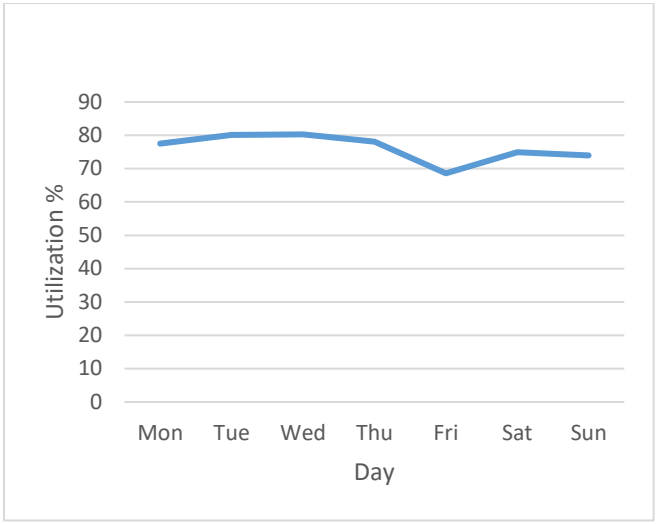


Fig 10. Linear graph of Average of WPUP in Summer

As shown in Fig10, the maximum link utilization of using the internet was on the middle of the week and reaches 80%, where the minimum link utilization during weekend, and it is around 67 %.

V. Conclusion

The internet international traffic data for LITC company was gathered for long period of time. This data was used to model the daily and weekly usage pattern of internet. The gathered data was obtained from Deep Packet Inspection (DPI) tool installed on the LITC network. The obtained daily pattern shows that there are two peaks of usage around 2:30PM and 12:00AM. Also, it has been shown that there is no difference in the daily usage patterns over the week in all seasons except summer, where the weekend usage was lower slightly than working days in Summer. The results were compared with other patterns obtained from different works and places. The link utilization was calculated for the seven weekdays. The value was just below 70% of the total link capacity all over the year.

References

[1] Tichi Zhang, "Residential Network Traffic and User Behavior Analysis", in KTH - Royal Institute of Technology, November 2010.

[2] B.Stewart, "Internet History", white paper, available in http://www.livinginternet.com/i/ii_summary.htm, April 2010.

[3] Internet World Stats, http://www.internetworldstats.com/stats.htm, retrieved April 1st, 2010.

[4] Argha Ghosh, Dr. A. Senthilrajan, "Classifying Network Traffic Using DPI And DFI", in International Journal of Scientific & Technology Research, volume 8, issue 11, November 2019.

[5] Pravesh Kumar Thakur, "Deep Packet Inspection", White Paper, Available: https://pdf4pro.com/amp/view/deep-packet-inspection-46f647.html 26.12.2019

[6] Safa Alkateb "White Paper: 5 Things You Need to Know About Deep Packet Inspection (DPI)" Available: https://docplayer.net/7150123-5-things-you-need-to-know-about-deep-packet-inspection-dpi.html

[7] "White paper on "Deep Packet Inspection" Available: http://tec.gov.in/pdf/Studypaper/White%20paper%20on%20DPI.pdf

[8] Po-Ching Lin, Ying-Dar Lin, Tsern-Huei Lee and, YuanCheng Lai "Using String Matching for Deep Packet Inspection", Available: https://ieeexplore.ieee.org/document/4488244

[9] Nate Anderson, “Data from ISPs in Europe and North America shows a big difference in peak”, white paper available: https://arstechnica.com/tech-policy/2009/09/does-less-evening-internet-mean-europeans-lead-better-lives/, January 2009.

Diagnosis and Surveillance of Covid-19 Pandemic Based on 3D Integral Images Technique

Mahmoud Geat Eljadid
Dept. of Software Engineering, Faculty of Information Technology,
Tripoli University,
Tripoli, Libya, P.O. Box: 13086
E-mail: meljdid@hotmail.com
E-mail: M.Eljadid@uot.edu.ly

Abstract—3D integral imaging is a true 3D imaging technology. It offers the simplest form that is capable of recording and replaying the true light field 3D scene in the form of a planar intensity distribution, by employing microlens array. There is a new world health crisis threatening the humanity with the spread of Covid-19 (Coronavirus Disease-2019). The Covid-19 belongs to a family of viruses that may cause various symptoms such as pneumonia, fever, breathing difficulty, and lung infection. Real images of Covid-19 patients confirmed by computed tomography CT were used to segment areas of increased attenuation in the lungs, all compatible with ground glass opacities and consolidations. This paper describes a new method to generate an indicative Covid-19 3D integral image models. The method is based on a Multiprocessor Ray Tracing System including Philips viewer, 3D slicer software and unidirectional camera. Experimental results are extremely satisfactory and for the first time it is proved that 3D integral images Covid-19 models are generated through Multiprocessor ray tracing system in order to deep monitoring and visualization to could be aid diagnosis in the absence of RT-PCR kits as demonstrated. A new file format content is created as well.

Keywords—computer graphics, 3D integral images, multiprocessor ray tracing, Covid-19 pandemic, medical images processing & visualization

I. INTRODUCTION

The main idea behind this novel algorithm is to utilize the previous work [1-5] to handle the generation of a covid-19 3D integral image a sequence of frames , and the exploit the new file format to prevent the distortion of the images.

Medical applications 3D endoscopy and tomography for direct observation and diagnosis of Covid-19 is now certain and available application. Creation of real-time 3D covid-19 imaging system has been now achieved to open application.

Existing Radiologic Representation of pathology is limited by its three-dimensional representation on a two-dimensional screen. [2][6].

In this new developed method, Philips Dicom viewer, 3D slicer packages and Multiprocessor ray tracing system are used in order to generate a Covid-19 3D integral images model. As a result, the new method is accomplished by adding a new programming functions in C, C++, Java and Mathlap to adapted Multiprocessor ray tracing system *"Tachyon"* software package, in order to easily diagnoses and observation of the Covid-19.

The new technique for confirmed case, reconstruction in motion and maximal intensity perspective projection do create Covid-19 3D integral images model. 3D reconstruction provides multiple projections with a preview of the surgical field and study of lesion characteristics, which can help achieve faster and safer surgery.

II. ADVANCED INTEGRAL IMAGES SYSTEMES

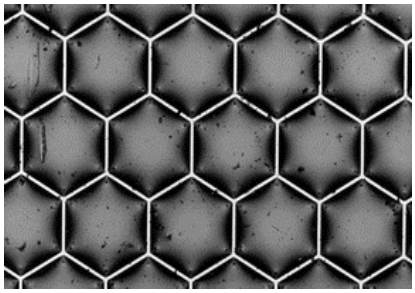
It is possible to capture 3D integral images electronically using a commercially available CCD array [22-38]. This form of capture requires a high resolution CCD together with specialised optical components to record the micro-images fields produced by precision micro-optics. The object/scene is recorded by a CCD placed behind the recording microlens array through a rectangular aperture. The aperture greatly affects the characteristics of the micro-images recorded. Since each micro-image is an image of the object seen through the aperture independently, its shape and size is determined by the aperture. If the field of a sub-image is fully covered by the image, it is said to be *fully-filled*, otherwise it is said to be *under-filled* or *over-filled*.

The system will record live images in a regular block pixel pattern. The planar intensity distribution representing an integral image is comprised of 2D array of $M \times M$ sub-images due to the structure of the microlens array used in the capture and replay. Sections of such typical lens array are illustrated in Figure 1. Different configuration patterns can be used in the design and manufacturing of microlens arrays as shown in figure 1. The *packing density* or *fill factor* is an important design criterion. The hexagonal arrangement of element microlenses has a higher capacity of the lens grid, and the hexagonal element shape can lead to 100% packing density without dead space [7]. These properties of the hexagonal microlens array make it a good choice for OII.

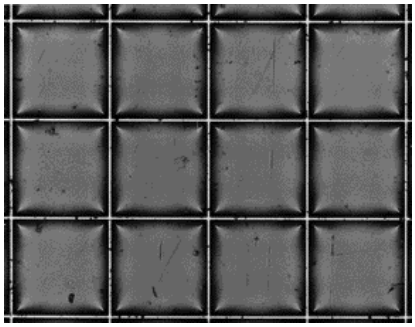
The resulting 3D images are termed Omnidirection Integral Images (OII) and have parallax in all directions. The rectangular aperture at the front of the camera and the regular structure of the hexagonal microlenses array used in the hexagonal grid (recording microlens array) gives rise to a

regular 'brick structure' in the intensity distribution as illustrated in Figure 2. Unidirectional integral images (UII) are obtained by using a special case of the integral 3D imaging system where 1D cylindrical microlens array is used for capture and replay instead of a 2D array of microlenses. A section of a cylindrical lens array is shown in Figure 3. The resulting images contain parallax in the horizontal direction only. Figure 4(a) shows an electronically captured unidirectional integral 3D image and Figure 4(b) shows a magnified section of the image. The *M* vertically running bands present in the planar intensity distribution captured by the integral 3D camera are due to the regular structure of the 1D cylindrical microlens array used in the capture process.

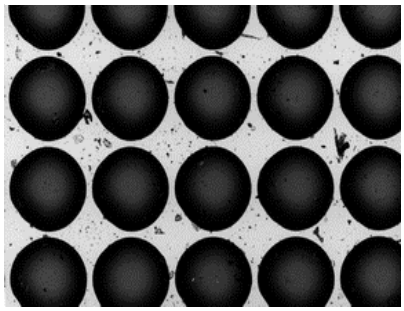
The replay of the 3D Integral images is achieved by placing a microlens array on the top of the recoded planar intensity distributions. The microlens array has to match exactly the structure of the planar intensity distribution.



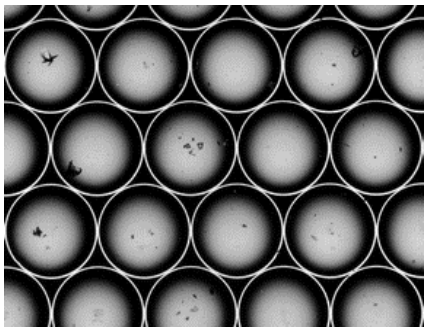
(a) hexagonal microlens array.



(b) Square microlens array.

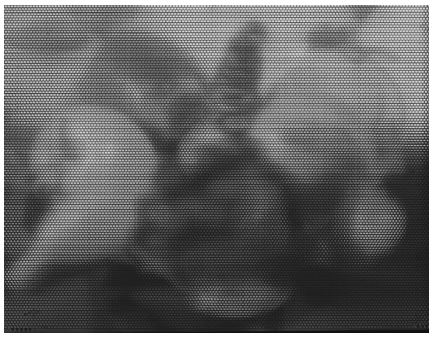


(a) Circular microlens array.
Orthogonal grid arrangement.

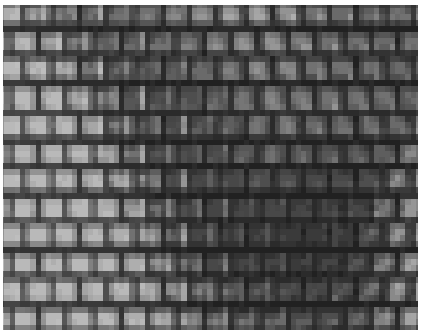


(b) Circular microlens array.
Hexagonal grid arrangement.

Figure 1: Enlarged part of microlens arrays using different configuration [8].



(a)



(b)

Figure 2: (a) Example of the nature of sub-image field. (b) magnified section [8].

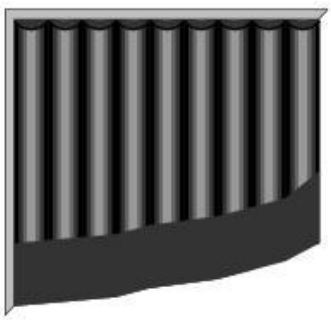
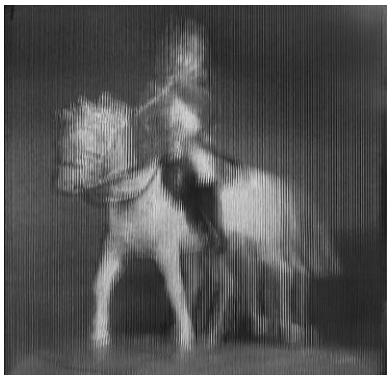
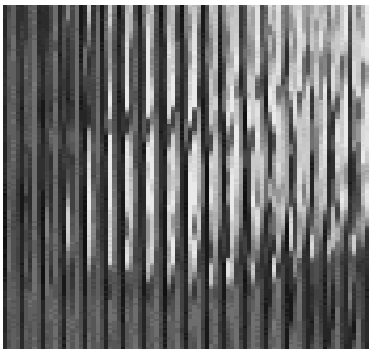


Figure 3: Diagrammatic representation of the lens array [9].



(a)



(b)

Figure 4. An electronically captured unidirectional integral image (a) Full (b) Magnification [8].

III. CAMERA MODEL IMPLICATIONS

Due to the nature of the recording process of integral imaging, many changes the camera model used in standard ray tracing. For lenticular sheets, each lens acts like a cylindrical camera. A strip of pixels is associated with each lens forming a sub-image. Each lens records a sub-image of the scene from a different angle as shown in the Figures 5 and 6. For micro-lenses arrays each lens acts like a square or a hexagonal camera depending on the structure of the lenses, as shown in Figure 7. In the lateral cross section of the lenticular or the micro- lenses, a pinhole model is used. In the case of lenticular sheets, the pinhole forms a straight line parallel to the axis of the cylindrical lens in the vertical direction. For each pixel, a primary ray is spawned. The recording path of the primary ray draws a straight line going forward towards the image plane and backward away from the image plane. Similar primary rays of neighbouring lenses are spawned to similar directions parallel to each other. Therefore highly correlated sub-images are produced which is a property of integral imaging.

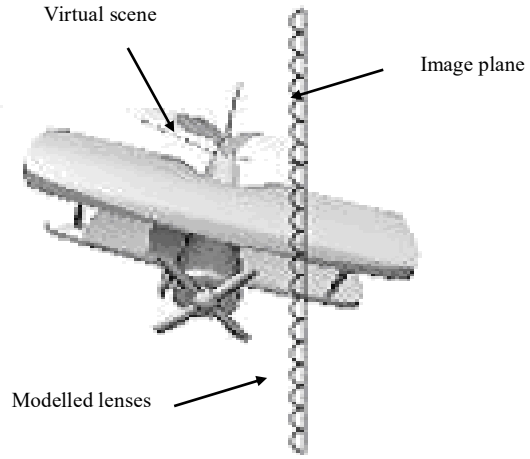


Figure 5: Top view of the modelled optical system of integral imaging [8].

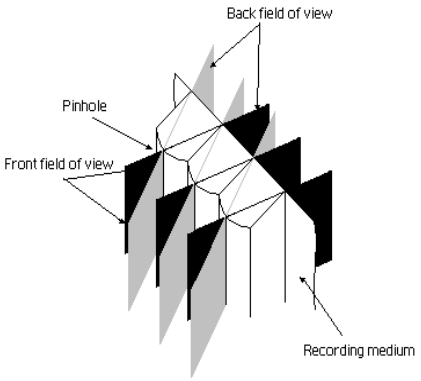


Figure 6: Lenticular sheet model in integral ray tracer [8].

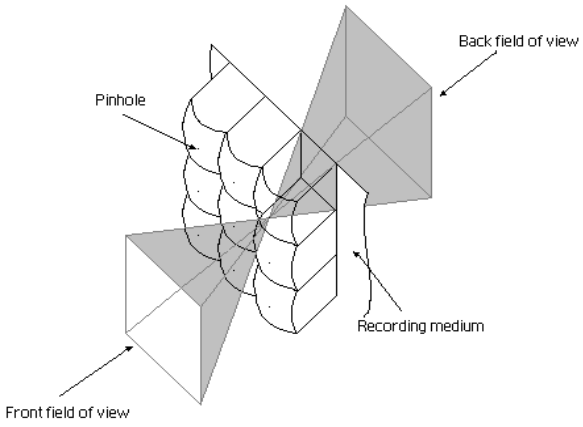


Figure 7: Micro-lens array in integral ray tracing [8].

The structure of the lenses and the camera model in the integral ray tracing affects the way primary rays are spawned as well as the spatial coherence among them.

IV. Computer Generation of Integral Imaging Using Ray Tracing

All the described properties of ray tracing can be migrated to three-dimensional integral imaging and opens the door to computer graphics applications that facilitate integral imaging displays. Integral imaging increases the perception of the photo-realism in ray tracing for the observer. Computer generation of integral imaging has been reported in several literatures [10-14]. Computer generation of integral imaging is very useful where an integral image can be replayed using a LCD monitor by overlaying it with a lenticular sheet.[19] modelled the optical system of integral imaging and applied it inside a ray tracing renderer. This new renderer is represented in the work by the term integral ray tracing (IRT) [32][37]. For both lenticular sheets and microlens arrays, each cylindrical lens or microlens acts like a separate camera. The virtual scene is straddling the modelled lenticular sheet as well as the image plane as shown in Figure 3.3. In recent years several research groups have proposed similar techniques for generating synthetic integral images [10] [12] [14-18][32-38]. However, most of the work concentrated on reproducing the various physical setups to using computer generation software packages such as *TACHYON* [19] and *POVRAY* [20].

V. Analysis Of Covid-19 3d Integral Images Models

The following described method leads to use 3D integral images for direct observation of Covid-19. Philips viewer software is used to read real CT scan sequence of images *113* frames or CT images for confirmed Covid-19 case as shown in figures 10 and 11.

This phase of the method is to load the images to a 3D Slicer software, then use the segmentation and volume rendering. Ultimately, the Covid-19 model is exported as **.obj* file. see Figure 8. There is a problem that has been encountered with importing the **.obj* file to an adapted multiprocessor ray tracing system is how to except the proper format of the new file.

The answer is to add unidirectional camera parameters to the new scene description file see Figure 9 and [5]. Eventually, The Covid-19, 3D integral images are generated see Figures 13, 14 and 17.

```
# 3D Slicer output. SPACE=LPS
mtllib right lung.mtl
v -30.9241 67.18 -938.605
v -30.2498 67.1033 -938.385
v -32.5224 67.9554 -938.655
 v -31.7196 67.6447 -938.791
vn 0.626329 -0.77382 0.0944195
vn -0.401724 0.914948 0.0385832
vn -0.496852 0.866905 0.0401653
f 3361//3361 4949//4949 4950//4950
f 4950//4950 4949//4949 6704//6704
f 4950//4950 6704//6704 6705//6705
……
……
```

Figure 8: **. Obj* file format.

```
3D-Model
begin_scene
resolution 1024 768
camera
FOCALLENGTH  6.8
LENSPITCH  2.116667
LENSPIXELS  9
APERTUREDISTANCE  10.0
SIZE  38.0
ZOOM  1.0
aspectratio 1.0
antialiasing 0
raydepth  2
CENTER 4.86 7.2 5.4
VIEWDIR -0.475149 -0.703924 -0.527943
UPDIR -0.29537 -0.437585 0.84928
END_CAMERA
```

Figure 9: unidirectional camera 3D integral images parameters.

VI. Phlips Dicom Viewer Software

The Philips DICOM Viewer is an application that is used to open DICOM data The application consists of a series selector and a series viewer and supports basic image viewing operations like playing movies and adjusting image settings. The Philips DICOM Viewer is a read-only application [39] as shown in Figure 10.

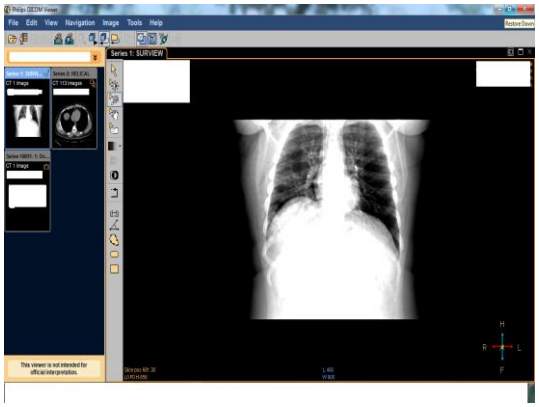
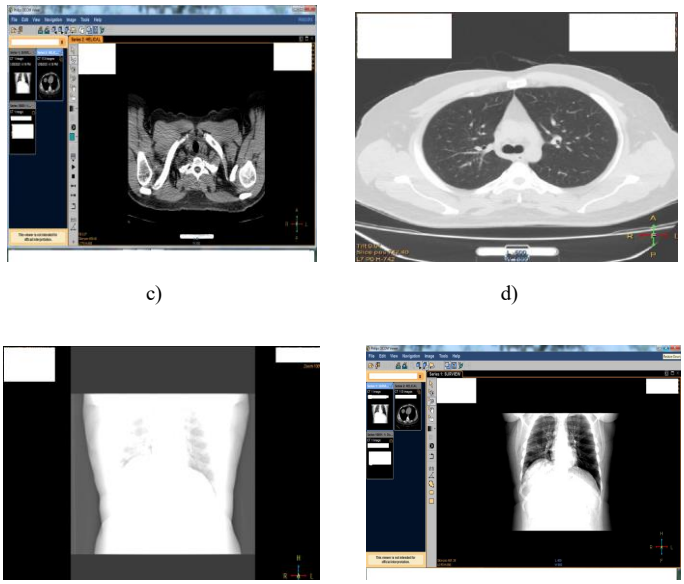


Figure 10: Screenshot of Philips multimodality DICOM viewer.

a)

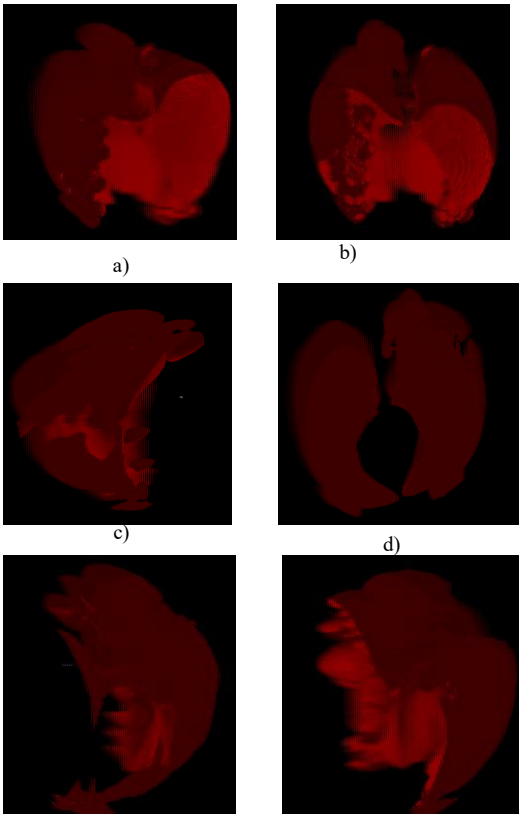
b)



c) d)

e) f)

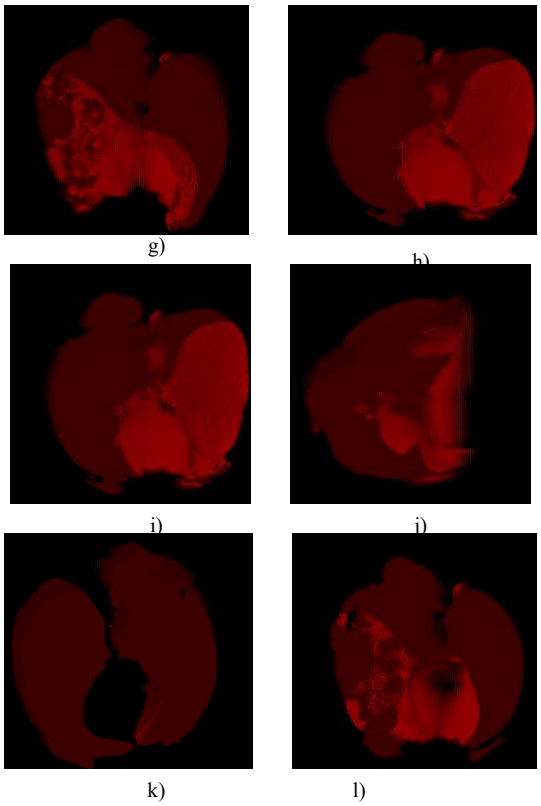
Figure 11 (a-f) : CT images of Covid-19 obtained from a clinics and hospitals are displayed by Philips viewer software.



a) b)

c) d)

e) f)

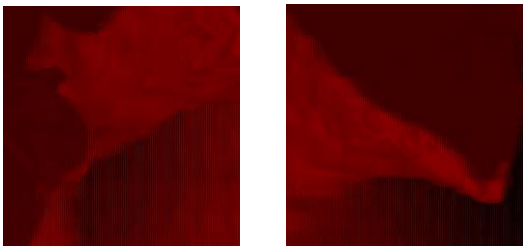


g) h)

i) j)

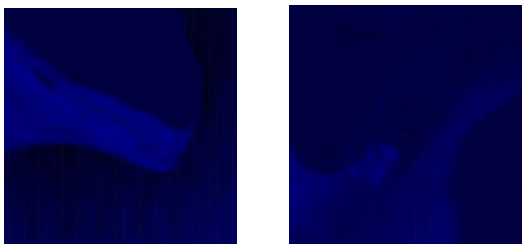
k) l)

Figure 12 (a-l): Covid-19 3D integral image frames of lung.



a) b)

Figure 13 (a,b): Zoom in view of Covid-19 3D integral images, camera *SIZE* 30.0.

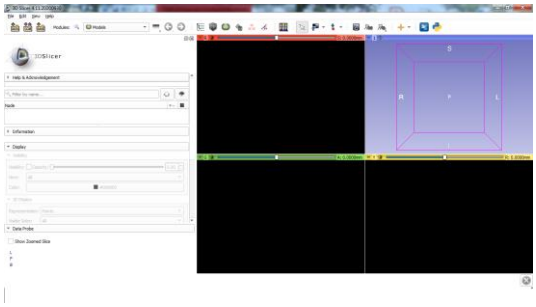


a) b)

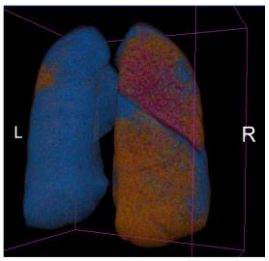
Figure 14 (a,b): Zoom in view of Covid-19 3D integral images camera *SIZE* 38.0. Different positions of camera

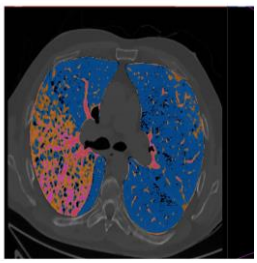
VII. 3D SLICER DICOM SOFTWARE

A software application for visualization and analysis of medical image computing data sets. All commonly used data sets are supported, such as images, segmentations, surfaces, annotations, transformations, etc., in 2D, 3D, and 4D. Visualization is available on desktop and in virtual reality. Analysis includes segmentation, registration, and various quantifications [40], as shown in Figures 15 and 16.



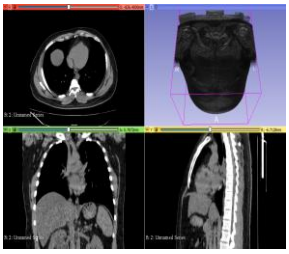
a)



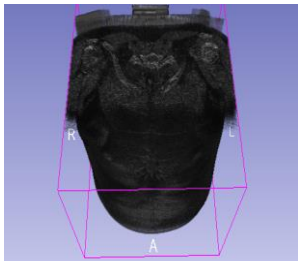


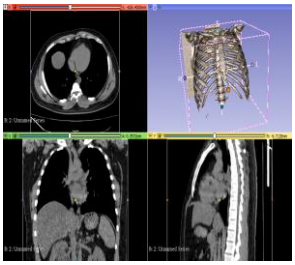
b) c)

Figure 15 (a-c): Cross-platform open-resource medical image processing and visualization system of 3D slicer [40].

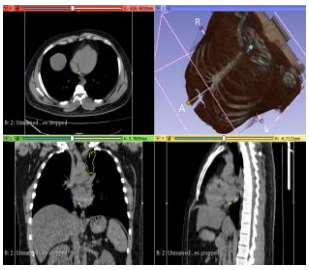


a) b)

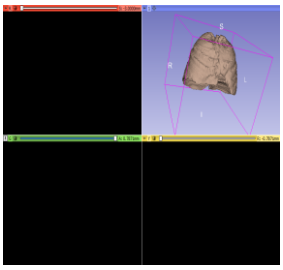




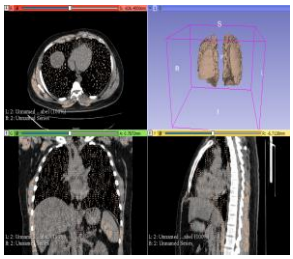
c) d)



e)



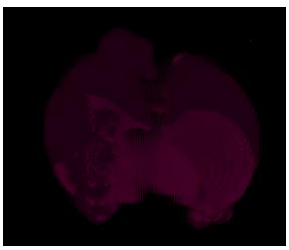
f)



g)

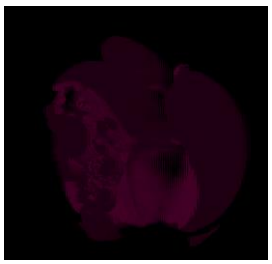
h)

Figure 16 (a-h) : CT images of Covid-19 are displayed by 3D slicer software.



a)

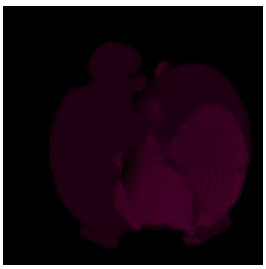
b)



c)

d)

e)



f)

Figure 17 (a-f):different sequence of Covid-19 3D integral image frames of lung.

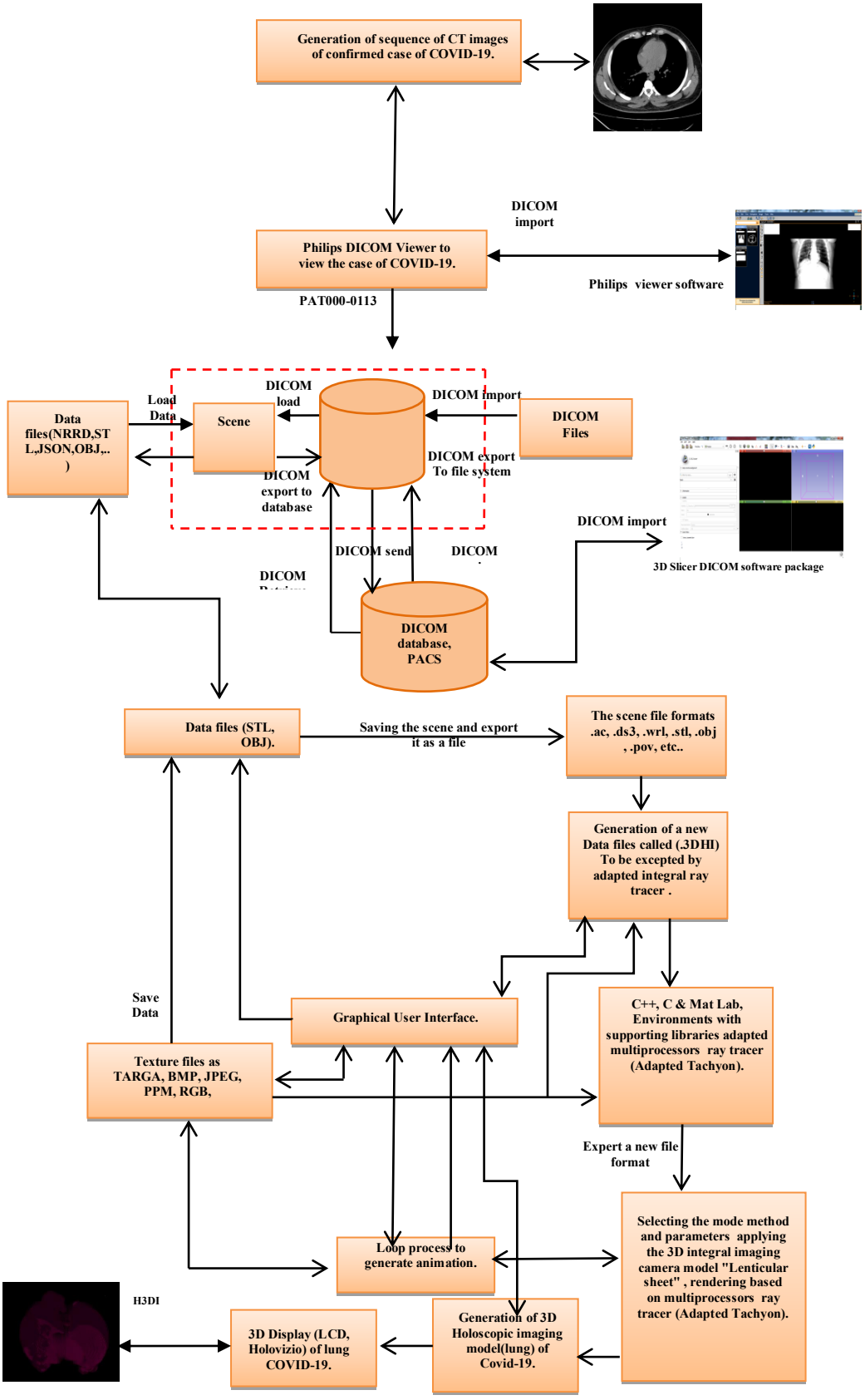


Figure 18:Flowchart Of Novel Generation Covid-19 3D integral images.

VIII. Experimental And Result

The results are extremely satisfactory and for the first time it is proved that Covid-19 images can be modeled to 3D integral images based on adapted parallel ray tracing system. In this paper a unidirectional integral images camera model is adopted. A new file format is created. An example of rendered scenes such a lung confirmed case of Covid-19 using the multiprocessor integral imaging ray tracer is shown in Figures 13, 14, 17 and 18.

IX. Conclusion

This paper presents new method to allow a computer generation of Covid-19 3D integral images modeling that would lead to easily diagnose and surveillance this virus. In this report, the CT images are loaded and then viewed by commercially available software tools such as Philips viewer and 3D slicer. The experiment was conducted on real CT scan data of Covid-19 case provided by clinics and hospitals achieving promising results. To my knowledge the first time this has been achieved. For this study the Adapted Multiprocessor ray tracing system Tachyon has been used as render. A new Covid-19 3D integral images description scene file format is generated.

X. References

[1] M. G. Eljadid, A. Aggoun, O. H. Youssef, "Computer Generated Content for 3D TV", *in proc of 3DTV conference,* Greece, 2007, .DOI:10.1109/3DTV.2007.4379381 ISBN: 978-1-4244-0722-4.

[2] Mahmoud G. Eljadid, A. Aggoun, " Medical 3D Integral Images Visualization in True Space" *Lecture Notes on Software Engineering, Vol. 4, No. 2, May 2016,* DOI: 10.7763/LNSE.2016.V4.229.

[3] Mahmoud G. Eljadid, Amar Aggoun, "3D Holoscopic Image Video Content Display on Volumetric Displays: The next generation 3D TV technology", *International Journal of Information Technology and Electrical Engineering*, Vol.7, No. 6, December 2018. ISSN:2306-708X.

[4] Mahmoud G. Eljadid, Amar Aggoun, "Computer Generation of 3D Integral Imaging Animations" Libyan International Conference on Electrical Engineering and Technologies, LICEET 2018, Tripoli-Libya 2018, LICEET13732018.

[5] Mahmoud G. Eljadid, Amar Aggoun, Osama H. Youssef Atallah, "New 3D Holoscopic Images Content Format" Libyan International Conference on Electrical Engineering and Technologies, LICEET 2018, Tripoli-Libya 2018, LICEET13732018.

[6] J. K. Makanjuola, A. Aggoun, M. Swash, Philipp C.R. Grange, B. Challacombe, P. Dasgupta, "3D-holoscopic imaging: A new dimension to enhance imaging in minimally invasive therapy in urology oncology," *Journal of Enduorology,* vol. 27, issue 5, May 2, 2013

[7] Akimakina, L.V. and Melnikova, N.V., "Basic Parameters and Properties of NIKFI, Hexagonal Lens Grid", Sov. J. Opt. Tech., vol. 35, pp.186-189, 1980.

[8] Wu, C., "*Depth measurement in integral images*," PhD Thesis, De Montfort University, 2003.

[9] Forman M., Aggoun A. and McCormick M., *"Compression of integral 3D pictures",* Fifth International Conference on Image Processing and its Applications, IEE conf., Pub. 410, pp. 584-588. 1995.

[10] S. Min *et al*, "Three-Dimensional Display System Based On Computer-Generated Integral Imaging," *Stereoscopic Display and Virtual Reality Systems VIII proc. Of the SPIE*, Vol. 4297, pp. 187-195, 2001.

[11] Halle, M. W. *et al*, "Fast Computer Graphics Rendering for Full Parallax Spatial Displays," *Practical Holography XI and Holographic Materials III Proc. Of the SPIE,* Vol.3011, 1997.

[12] Naemura T., T. Yoshida and H. Harashima, "3D Computer Graphics Based on Integral Photography," Optics express, Vol. 8(2), pp. 255-262, 2001.

[13] Grahm E. Milnthorp, "*Computer Generation of Integral Images using Interpolative Shading Techniques*" De Montfort University, PhD Thesis 2003.

[14] Athineos, S. Sgouros, N." Photorealistic integral photography using a ray-traced model of capturing optics" Journal of Electronic imaging Oct-Dec 2006/vol. 15(4).

[15] Ren, J., Aggoun, A., and McCormick, M., *'Computer generation of integral 3D images with maximum effective viewing angle' Journal of Electronic Imaging, Vol. 14, 2005.*

[16] Yang, R., Huang, X., and Chen, S., "Efficient Rendering of Integral images" Proceeding of SIGGRAPH 2005.

[17] Park, K. S., Min, S.W., Cho, Y., "Viewpoint vector rendering for efficient elemental image generation" IEICE Trans. Inf. & Syst., Vol. E90-D(1), pp. 233-241, 2007.

[18] Motoki, T., Isono H. and Yuyama I., *"Present status of three-dimensional television research", Proc. IEEE'83*, pp. 1009-1021, 1995.

[19] Parallel/Multiprocessor Ray Tracing Software. [Online]. Available jedi.-ks.-uiuc.-edu/~johns/-raytracer/

[20] http://www.povray.org/

[21] Yamazaki, K. Kamijo and S Fukuzumi: "Quantative evaluation of visual fatigue" Proc. Japan Display, pp 606-609. (1989).

[22] M. T.M. Lambooij W. A. IJsselsteijn, I. Heynderickx: "Visual Discomfort in Stereoscopic Displays: A Review" Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6490, (2007).

[23] Lippmann, G. '*Eppreuves Reversibles Donnat Durelief*', *J. Phys*. Paris 821 (1908).

[24] S Manolache, A Aggoun, M McCormick, N Davies, S Y Kung, "*Analytical model of a three-dimensional integral image recording system that uses circular and hexagonal based spherical surface microlenses"*, Journal of the Optical Society of America. pt A, 18,No.7, pp 1814-1821, Aug. 2001.

[25] A Aggoun: *'Pre-processing of Integral Images for 3D Displays'* IEEE Journal of Display Technology, Vol. 2. NO. 4, pp. 393-400, Dec. 2006.

[26] M McCormick, N Davies, A Aggoun: '*3D television and display systems using integral imaging*'. Invited paper, 'Photonics East' SPIE Conference on 3D Display Systems. Boston USA, Vol. 2864, pp. 51-59, July 2002.

[27] N Davies, M McCormick and Li Yang: "*Three-dimensional imaging systems: A new development*". Applied Optics. Vol 27, 4520, (1988).

[28] M McCormick, N Davies, A Aggoun, M Brewin: "*Resolution requirements for autostereoscopic full parallax 3D-TV*". International Broadcasting Conference, Amsterdam, Sept. 94. IEE Conference Publication No.397, (1994).

[29] Lippmann G, "La Photographic integrale" comtes Rendus, Academic des Sciences, 146, 446-451, 1908.

[30] Ives, H. E., "Optical Properties of a Lippmann Lenticulated Sheet," J. Opt. Soc. Am., Vol.20, pp.171-176, 1931

[31] M. Mccormick and N. Davies, "Full natural colour 3D optical models by integral imaging," Holographic Systems, Components and Applications, 4th International Conference on , vol., no., pp.237,242, 13-15 Sep 1993.

[32] M. G. Eljadid, "*3D content computer generation for volumetric displays,*" PhD Thesis, Brunel University West London, 2007.

[33] Contract no: IST-7-248420-STREP, Program FP7-ICT-2009-4. 1st Newsletter June 2010. Intermediate Dissemination Report. (20th April 2012). [Online]. Available: http://www.3dvivant.eu/.

[34] A. Aggoun, E. Tsekleves, D. Zarpalas, P. Daras, A. Dimou, L. Soares, P. Nunes, "Immersive 3D Holoscopic System", *IEEE Multimedia Magazine, Special Issue on 3D Imaging Techniques and Multimedia Applications,* Vol. 20, Issue 1, pp. 28-37, Jan-Mar 2013.

[35] M. G. Eljadid, A. Aggoun, O. H. Youssef," *Enhanced Techniques 3D Integral Images Video Computer Generated*" *Proceedings of the International conference on Computing Technology and Information Management,* Dubai, UAE, 2014. ISBN: 978-0-9891305-5-4 ©2014 SDIWC.

[36] M. G. Eljadid, A. Aggoun, O. H. Youssef," Enhanced Still 3D Integral Images Rendering Based on Multiprocessor Ray Tracing System" Journal of Image and Graphics, Volume 2, No.2, December 2014 doi: 10.12720/joig.2.2.117-122.
[37] O. H. Youssef, "Acceleration techniques for photo-realistic computer generated integral images," Ph.D Thesis, De Montfort University, 2004.
[38] O. H. Youssef and A. Aggoun, "Coherent grouping of pixels for faster shadow cache," in *Proc. 3rd Holoscopic Computer Graphics, 3DTV Conference*, Tampere, Finland, 8-9 June 2010.
[39] Philips Dicom viewer software package [online]. Available http:// https://www.philips.com/
[40] 3D Slicer Dicom software package . [online] Available https://www.slicer.org/

A Natural Language Processing Approach for the Digitalization of Roaming Agreement

Santiago Figueroa-Lorenzo
CEIT and University of Navarra
sfigueroa@ceit.es

Noureddin Sadawi
Softlight Technology
noureddin@softlight.tech

Ahmad Sghaier Omar
DarSys Tech
asghaier76@gmail.com

Mohamed Elshrif
Qatar Computing Research Institute
melshrif@hbku.edu.qa

***Abstract*—Roaming is the capability of a subscriber to have persistence of service in a visited mobile network managed by a different operator. To achieve this persistence of connectivity, it is necessary for operators to agree on technical, commercial and legal aspects in what is known as a Roaming Agreement (RA). Currently, RA negotiation uses asynchronous flow such as email or even regular mail. The main phase in building the RA is known as the drafting phase, which is a time-consuming process with limited assurance and transparency on handling the drafting events. To address the lack of transparency and reduce the drafting time, this paper proposes a framework that automates the process of drafting and negotiation of the RA and enables better accessibility. The proposed framework exploits a Natural Language Processing (NLP) engine as a starting point for the digitalization of the negotiation process towards transparent drafting of RAs. Extensive experiments on real RAs show the superiority of our proposed framework in drafting the RAs against the traditional drafting and negotiation process with prediction accuracy over 80.0%.**

***Index Terms*—Roaming Agreement, Natural Language Processing, Telecommunication**

I. Introduction

The roaming service maintains the persistent connectivity of subscribers in different networks and locations. Roaming describes the capability of a subscriber to access mobile services offered by the visited public mobile network (VPMN) through the home public mobile network (HPMN) when roaming outside the coverage range of the HPMN [1]. However, before ensuring persistent connectivity in VPMN the Mobile Network Operators (MNOs) must reach an agreement regarding the technical, commercial and legal relationships known as the Roaming Agreement (RA). Within the RA, three stages are possible to establish: *drafting*, *testing* and *implementation*. On the one hand, the *testing* and *implementation* phases are straightforward as they are focused on implementing what is agreed on in the *drafting* phase. On the other hand, the *drafting* phase is ambiguous and involves multiple iterations of the exchange of draft text and proposed articles between both parties, i.e. the MNOs.

Therefore, to reduce the exerted efforts and standardize the technical, commercial and legal aspects of the RA, the GSM Association broadly outlines the content of such RA in standardized form for its members [2]. In addition, many organizations attempt to unify RAs. For example, Rocco[1], which is a company specialized in telecommunication reports including RAs, provides a list of the most commonly used GSMA standards. It summarizes these standards as follows [3]: (1) AA.12 constitutes the permanent reference document; (2) AA.13 contains the common annexes with operational information (e.g., information on tap file, billing data, settlement procedure, customer care, fraud, etc.) and (3) AA.14 involves the individual annexes containing information about the operator (e.g., contact details of the roaming team, fraud team, IREG team, TADIG team, etc.).

While it is true that it is not mandatory to follow the standards proposed by the GSMA organization, according to authoritative voices in the field of negotiating RA *drafting*, most MNOs follow them strictly [4]. Therefore, the first point to consider in the RA *drafting* is how far it has deviated from the GSMA's proposed standards. Thus, during the *drafting* process of the agreement, the parties should analyze the sub-articles contained in the GSMA standard templates. This necessitates discretizing the bulk of the RA template and classifying each word/clause into existing classes as follows:

1) Specify the value of certain *variables* that are found in a certain text, such as dates, names of MNOs, locations and others.
2) Introduce certain *variations* in the articles/sub-articles, usually identified as part of the text with different paraphrasing.
3) Leave an article/sub-article as found in the template (RA draft); thereby establishing a *standard clause*.
4) Introduce completely new articles/sub-articles that respond to particular interests by constituting *customized texts*.

However, the *drafting* of a RA goes through a complex negotiation process in which, at present, the parties still use asynchronous flows such as e-mail or even regular mail to exchange the information. Thus, this traditional negotiation process has multiple drawbacks including lack of transparency, which can lead to violations of the RA by MNOs. In addition, this process is laborious and time-consuming. As experts point out, the entire process can take up to one month, depending on the responsiveness of the MNOs [4]. Therefore, it is necessary to provide a transparent digitalization system for RA *drafting* negotiations, which guarantees transparency and reduction of the negotiation time (which could take days or

[1]Rocco Group: https://www.rocco.group

even weeks). Hence, this work develops a framework that exploits the advancement in the Natural Language Processing (NLP) discipline and uses it as an engine to digitalize legal text. This NLP Engine constitutes the starting point for the digitalization of the negotiation process towards a transparent *drafting* of RAs. The proposed NLP approach analyzes articles and sub-articles of RAs by determining the existence of *variables*, *variations*, *standard clauses* and *customized texts*. To do so, the proposed NLP Engine relies on multiple NLP techniques such as Named Entity Recognition (NER) and Part of Speech (POS) tagging in the *drafting* of RAs. A good introduction to NLP can be found in [5].

The main contributions of this work can be summarized in the following points: (1) Automatic accurate extraction of text from RA PDF files (2) The use of a text similarity metric to identify various parts of the RA (3) Breaking down articles and sub-articles of RAs by determining the existence of *variables*, *variations*, *standard clauses* and *customized texts* (4) Taking advantage of Amazon's powerful NLP engine (i.e. Amazon Comprehend) to perform some NLP tasks in a speedy and highly accurate manner.

The remainder of this paper is organized as follows: we briefly review existing work in Section II and describe our methods in detail in Section III. We then present and discuss our results in Section V and follow it with conclusions and ideas for future work in the last section.

II. Related work

The existing work that is related to the RAs can be categorized into two main categories. The first category encompasses work which focuses on transparent digitalization of RAs [6]–[8], whereas the second category includes work that applies existing NLP techniques and text processing systems [9]–[11].

Both in the scientific literature and business environments, there are important approaches to RA digitalization. Thus, the work discussed in [6] proposes a dynamic RA between the Local 5G Operator and the MNO. The interaction between the two entities takes place through an Ethereum based platform. Despite the contributions of this work in the RA *implementation* phase, the fact of performing the implementation on the Ethereum network implies that the participating MNOs will bear the high cost of gas fees for the transactions. Additionally, although the authors justify the latency given that they repeated the experiment 100 times, the proof of work nature of the network implies drafting actions are likely to suffer congestion at any time. An approach that focuses uniquely on the billing of the services obtained as a result of the RA is reported in [7]. This agreement is incorporated as part of a chaincode of a Hyperledger Fabric Blockchain (HFB) network so the work contributes significantly to the digitalization process of the RA, allowing for a faster, more seamless process in which payments can be requested and obtained quickly due to less need for manual intervention. Although this work also makes an important contribution regarding the use of HFB in the field of telecommunications to address problems around Managing RAs, Inter-carrier Settlements, and Mobile Number Portability, its approach is one of review and proposal, however, it lacks an implementation section that would allow demonstrating the feasibility of the proposals. The contextualization of this system in the business environment is proposed by important MNOs such as Telefonica, Deutsche Telekom, and Vodafone which use blockchain for Roaming settlement within the framework of the RA between the parties [8]. This proposal is also of considerable value in the business environment, however, its focus relates only to the *implementation* phase of the RA and not to the *drafting* phase.

Additionally, the scientific literature addresses text processing systems based on NLP techniques in domains such as the judiciary domain. Thus, the approach introduced in [9] addresses the process of digitalization in the judicial sectors from archives of judicial records for which the authors have designed a text analysis tool that includes grammatical analysis of documents in English based on NLP techniques. The proposed article also makes an important contribution in terms of the design of the linguistic analysis based on NLP techniques. However, it lacks formal evaluation in terms of accuracy of NLP implementation. In addition, the work reported in [10] applies NLP-based processing techniques for information retrieval from spreadsheets and describes technologies for storing and retrieving database information. NLP techniques such as sentence tokenization, word tokenization, removing stopwords and lemmatization are part of the parsing stage of the work. The work represents a valuable contribution in terms of design and implementation of the tool, however, it lacks demonstration of the feasibility of use, as well as determination of the accuracy of the NLP techniques applied. Finally, authors of [11] propose a methodology for implementing sentiment analysis using some online services on the Amazon cloud. This system performs an audio-to-text transcription and then performs the processing of the obtained text. Although the value of the contribution lies in the use of NLP techniques from the Amazon cloud (mainly Amazon Comprehend for text analysis), the paper lacks a formal results section, therefore it is difficult to determine the accuracy of the proposed tool. One key issue in this work is the errors that can result after using audio-to-text transcription which can lead to further errors as processing continues.

Although these studies constitute a relevant part of related work, e.g., by integrating useful tools such as Amazon Comprehend or detailing the use of techniques such as tokenization, the scientific literature does not address scenarios for the telecommunications field and even less in the context of a transparent digitalization of the RA. Therefore, and to the best of our knowledge, we can affirm that our work introduces a topic with a high degree of novelty.

III. NLP Engine methodology

A. Detection and Comparison

In general, our design consists of the following two approaches (which we discuss in more detail in this section): *detection* and *comparison*. While *detection* has *symbols* as its basic processing unit, *comparison* has *sub-articles* as its basic

processing unit. *Detection* represents the capability to detect *variables* in a text file, i.e., the RA. For this purpose, *Amazon Comprehend* constitutes the enabling technology of the *detection* approach, since it is a service that uses NLP techniques to extract insights about the content of text documents by recognizing entities, key phrases, language, sentiments, and other common elements in a text [12].

Comparison represents the capability to find *similarities* and differences between the sub-articles present in the RA concerning the sub-articles present in the GSMA standard template. Our comparison approach is based on *text similarity* [13], since it is a resource commonly used for pattern classification, clustering, and information retrieval problems [14]. We employ *Jaccard's similarity*, which is defined as the size of the intersection divided by the size of the union of two sets [15]. As a result of the *comparison*, it is determined that while an almost total coincidence between texts at the sub-article level represents a *standard clause*, an almost null coincidence between texts (or simply the non-existence of a sub-article of the GSMA standard template in the RA) represents a *customized text*. Thus, the intermediate case is represented by the *variation* in which there is a high coincidence between sub-articles, and the existing differences are given by the presence of *variables* such as the commercial names of MNOs and the start date of the RA. The next section integrates tools, NLP techniques, and text processing techniques as part of the designed methodology.

B. Designed Methodology

The flow chart in Fig. 1 is a general scheme of the methodology designed for the NLP Engine, which will be explained in detail below.

The starting point of the methodology is the decompression and content extraction of PDF files (each PDF file contains a RA). The file name is used as one of the identifiers of the output file. The next step is to find similarities in document texts so that headers and footers are detected to avoid undesired characters. For that purpose, the mechanism used is the detection of different font sizes and weights. Once the PDF format has been converted to plain text and the removal of headers, footers, and undesired characters (e.g., spaces, endlines) is performed (in what we call a parsing step), it is now possible to perform both *detection* and *comparison*.

The *detection* phase includes a requirement associated with the Amazon Comprehend tool regarding the number of symbols to be sent via REST API (it accepts text of a small size). For this reason, the text is segmented into pieces of 100 characters. For each character chunk the API outputs entities, key phrases and sentiment in JavaScript Object Notation (JSON) format [12], which must be grouped and further processed. The designed logic collects the most important fields such as: 'BeginOffset', 'EndOffset', 'Score', 'Text', 'Type', 'PartOfSpeech', and 'Tag'. In turn, it adds to these the 'Frequency' field that indicates the number of appearances of entities with the same name, i.e., the same 'Text' field. Once this filtering has been done, entities are sorted according

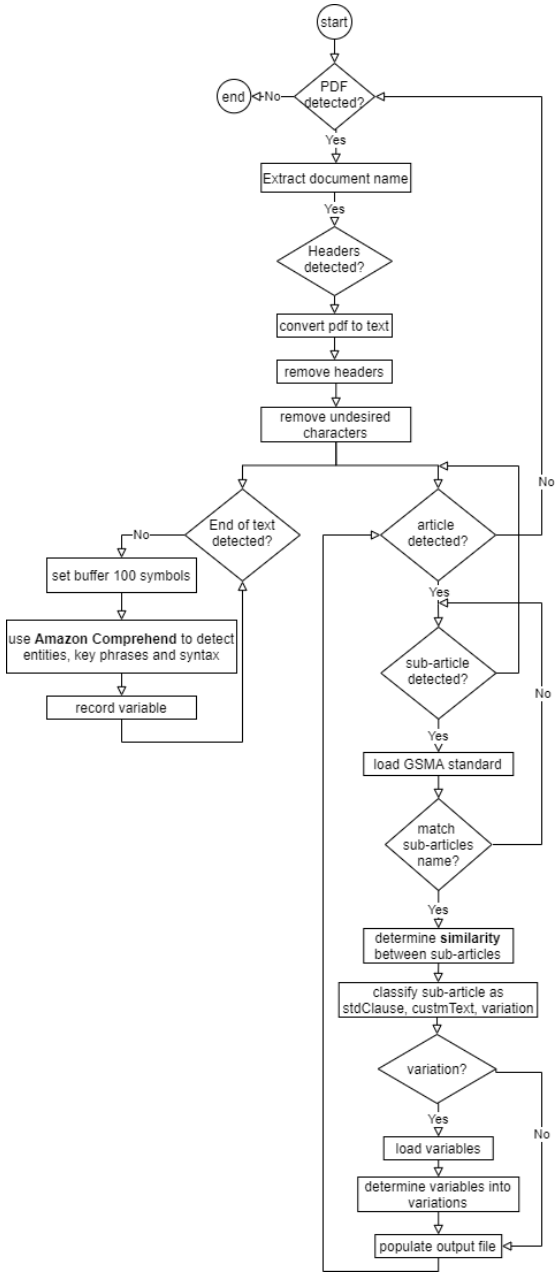


Fig. 1. Overview of the designed methodology.

to a logic that defines their relevance degree. The designed logic is particularized according to the type of *variable* to be detected. For instance, for the name of the MNO, the 'Score', 'Frequency', and 'BeginOffset' are prioritized, in addition to verifying the number of occurrences within the key phrases. The detected *variables* are stored to be used later to determine which ones are part of the *variations* or *customized texts*.

The *comparison* phase begins by dividing the Roaming Agreement first by articles and then by sub-articles. To compare the divided sub-articles, the NLP engine should load the GSMA standard template sub-articles as an additional file. *Jaccard's similarity* associates a score to the sub-article

ID. Thus, a sub-article is considered a *standard clause* when the score assigned to it is at least 0.85 (score $\geq$ 0.85). In addition, a sub-article is considered a *custom text* when the score assigned to it is 0.15 at maximum (score $\leq$ 0.15). Finally, a sub-article is considered a *variation* if the score assigned to it is between 0.85 and 0.15 (0.15 $<$ score $<$ 0.85). The NLP engine then populates the output file. For this purpose, it inspects whether the sub-articles have been tagged as variation or custom texts. If this is the case, it proceeds to identify the existing variables.

IV. System Implementation

Fig. 2 shows the overall architecture of the NLP engine integrated into a docker infrastructure that includes three parts. The input files include the RA text file as well as the GSMA standard templates. The processing layer includes the logic associated with the NLP engine, i.e., the implementation of the designed methodology. The output is a JSON file populated with the classification of sub-articles as *standard clauses*, *customized texts* and *variations*. In addition, each article includes the set of *variables* it contains and each sub-article contains the specific *variable* detected as long as it has been classified as *variation* or *customized texts*.

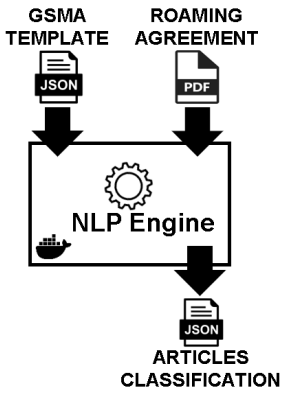


Fig. 2. NLP Engine overall architecture.

The input and output files are accessed by the container through Docker volumes. In addition, the NLP Engine has been developed as a Python v3.8 based library and therefore must be imported and run from an entry point. Although the NLP Engine constitutes a library, it must in turn, integrate other libraries such as boto3 [16] and PyMuPDF [17]. Boto3 is a Software Development Kit (SDK) for Python, created and supported by Amazon Web Services (AWS). The NLP Engine imports the specific service associated with Amazon Comprehend. PyMuPDF constitutes another library imported into the NLP Engine to be used for headers and footers detection. Therefore, the docker image built for the NLP Engine must not only install the NLP Engine library but must also install the libraries it imports. The *similarity* has been implemented in a pluggable way, therefore at code level, it is easy to change the *Jaccard's similarity* with another one, e.g., cosine similarity [15].

V. Discussion of Results

Our NLP Engine architecture receives PDF files as input. Because PDF files normally consist of unstructured text, undesired characters may remain despite text parsing. Therefore, it is mandatory to determine the *accuracy* of the results obtained once the output JSON file has been populated. For this purpose, two types of approaches have been conducted. Namely, a simple inspection at the sub-article level and a verification based on symbol comparison. The tests have been performed on two real RA samples from the mobile operators Proximus and Orange [18]. Each experiment is described below and the results obtained are discussed in the same section.

A. Accuracy based on a simple inspection at sub-article level

The first accuracy analysis consists of visually determining (i.e. human-eye inspection) whether each sub-article of the RA constitutes a *variation*, a *standard clause* or a *customized text* concerning the GSMA standard template. The results obtained are then compared with the values populated in the NLP Engine output file, for the same sub-article. Considering that the results obtained from the simple inspection represent the observations and the values collected from the sub-articles classification file represent the predicted values, the following confusion matrices allow us to analyze the results for each sample RA. The confusion matrix in Table I shows satisfactory results for the NLP Engine's predictive capability in terms of accuracy (the NLP Engine correctly classifies 61 out of 72 analyzed sub-articles as *standard clauses*, *variations* or *customized texts* for the Proximus sample RA). Proof of this is the fact that the main diagonal of the matrix contains the highest number of True classifications.

TABLE I
Confusion Matrix for Proximus Roaming Agreement.

n = 72	***stdClause***	***variation***	***customText***
stdClause	21	3	1
variation	2	35	4
customText	0	1	5

TABLE II
Confusion Matrix for Orange Roaming Agreement.

n = 86	***stdClause***	***variation***	***customText***
stdClause	14	0	1
variation	1	8	0
customText	4	1	57

Table II shows satisfactory results for the NLP Engine's predictive capability in terms of accuracy (the NLP Engine correctly classifies 79 out of the 86 sub-articles as *standard clauses*, *variations*, or *customized texts* for the Orange sample RA). Using the same evaluation criteria, it is observed how the main diagonal of the confusion matrix in Table II has the highest number of True Classifications. Table III also summarizes the accuracy of the NLP Engine's sub-article classification capability in percentage terms as *standard clauses*, *variations* and *customized texts*.

TABLE III
SUMMARY OF ACCURACY DETERMINATION FOR SIMPLE INSPECTION AT SUB-ARTICLE LEVEL.

	Proximus	***Orange***
stdClause	80.9%	92.9%
variation	82.9%	87.5%
customText	80.0%	91.2%

Beyond the fact that the results themselves can be considered acceptable, the other behavior to be highlighted in Table III is the greater classification capacity into *standard clauses*, *variations* and *customized texts* of the NLP Engine in one document concerning another.

B. Accuracy determination based on symbol comparison

The second accuracy analysis involves establishing a comparison between the sub-articles populated in the output file concerning the sub-articles existing in the input file containing the real RAs. This comparison is performed at the symbol level, considering the order of appearance of each symbol. For that purpose, the text comparison tool Countwordsfree [19] is used by manually copying sub-article by sub-article. For each sub-article, the following values are determined:

1) Common percentage of symbols between compared sub-articles.
2) Difference in percentage of symbols between compared sub-articles.
3) Common symbols between compared sub-articles.
4) Difference symbols between compared sub-articles.

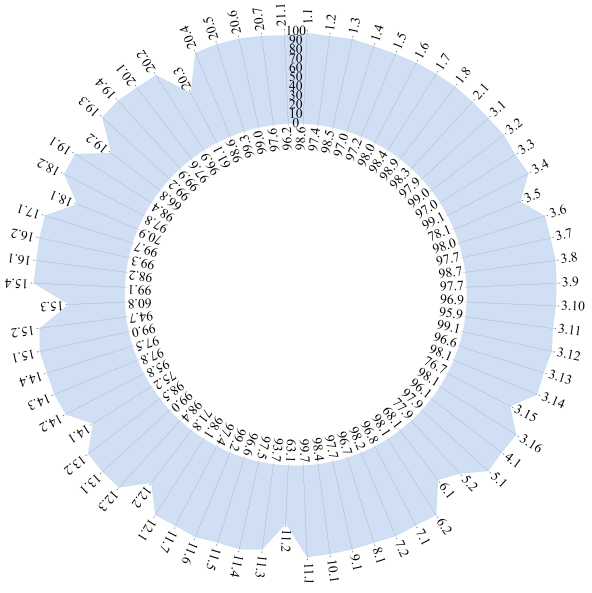


Fig. 3. Common percentage of symbols for Proximus Roaming Agreement.

The illustrated radar charts allow us to analyze the results for each RA. The radar chart contains in the inner radius the common percentage of symbols between compared sub-articles and in the outer radius the identifier of the compared sub-article, i.e., the sub-articles contained in the output file. Fig. 3 shows a comparison in terms of the common percentage of symbols between compared sub-articles for the Proximus sample RA. The conclusion that can be reached by simple inspection is that only 11 out of the 72, i.e., 84.7% of the analyzed sub-articles are affected, mostly by the introduction of undesired characters when the output file was populated.

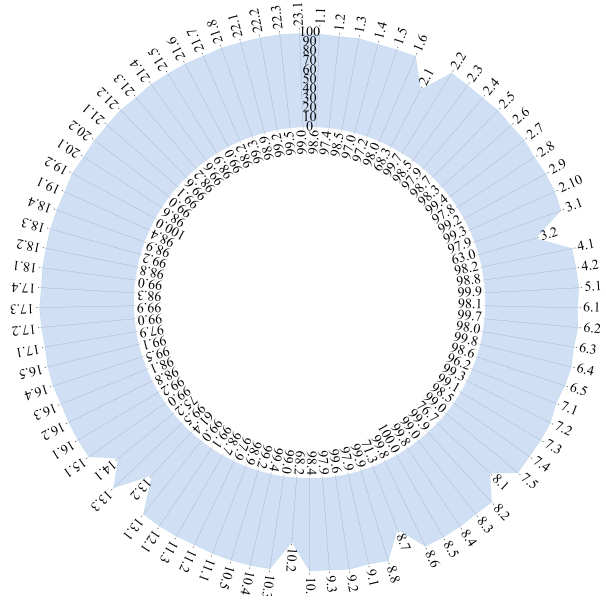


Fig. 4. NLP Engine overall architecture for Orange Roaming Agreement.

Similarly, Fig. 4 shows a comparison in terms of the common percentage of symbols between the sub-articles compared for the Orange sample RA. In this case, only 7 of the 86, i.e., 91.9% of the analyzed sub-articles are affected by the introduction of undesired characters when the output file is populated, thus improving the results obtained regarding the RA Proximus sample.

VI. CONCLUSIONS

The designed NLP-based methodology presented in this paper digitizes the RA by accurately extracting its contents and classifying its sub-articles as *standard clauses*, *variations* and *customized texts*. The test phase allows to evaluate the results obtained from two types of experiments conducted to measure the accuracy to establish that on the one hand, the accuracy determination based on a simple inspection at sub-article level determines for the first sample roaming agreement, and accuracy of 80.9% in the classification of *standard clauses*, an accuracy of 82.9% in the classification of *variations* and accuracy of 80% in the classification of *customized texts*. These results are improved for the second sample RA with an accuracy of 92.9% in the classification of *standard clauses*, an accuracy of 87.5% in the classification of *variations* and an accuracy of 91.2% in the classification of *custom texts*. On the other hand, the determination of the accuracy based on the comparison of symbols determines for the first roaming agreement sample, an accuracy of 84.7% of the total of 72

analyzed sub-articles with a common percentage of favorable symbols. This result is also improved for the second roaming agreement sample with an accuracy of 91.9% of the total of 86 sub-articles analyzed with a common percentage of favorable symbols. Therefore, the results demonstrate the feasibility of applying the proposed methodology. As part of a process of continuous improvement of the designed methodology within future research lines we aim to conduct other accuracy tests applying other text similarity types, such as cosine similarity, to improve the results obtained.

The proposed NLP Engine represents the starting point (i.e. the first module) of a project that primarily aims to transform the current Telecommunication Roaming Agreement drafting and negotiation process into a digitalized version based on the transparency promoted by blockchain technology. Future research can include the design, development, and evaluation of the rest of the sub-modules that mainly include the smart contracts that automate the negotiation process, as well as the integration with the developed NLP Engine.

VII. Acknowledgment

This research was funded by Linux Foundation Mentorship Program through the project: "The Use of NLP and DLT to Enable the Digitalization of Telecom Roaming Agreements" with project identifier d8a154c6-41fb-4733-b3c8-df37796e7fa3.

References

[1] I. Tanaka, "VoLTE Roaming and Interconnection Standard Technology," *NTT DOCOMO Technical Journal*, vol. 15, pp. 37–41, 2013. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/rd/technical_journal/bn/vol15_2/vol15_2_037en.pdf.

[2] R. Ferwerda, M. Bayings, M. van der Kam, and R. Bekkers, "Advancing E-roaming in Europe: Towards a single "language" for the European charging infrastructure," *World Electric Vehicle Journal*, vol. 9, no. 4, pp. 1–15, 2018, ISSN: 20326653. DOI: 10.3390/wevj9040050.

[3] R. Research, *The International Roaming Agreement*, 2017. [Online]. Available: https://www.roccoresearch.com/portfolio-items/the-roaming-agreement/ (visited on 08/24/2021).

[4] ——, *What is ROAMING HUBBING?* 2017. [Online]. Available: https://www.roccoresearch.com/portfolio-items/roaming-hubbing/ (visited on 08/24/2021).

[5] D. Jurafsky and J. H. Martin, *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, N.J.: Pearson Prentice Hall, 2009. [Online]. Available: http://www.amazon.com/Speech-Language-Processing-2nd-Edition/dp/0131873210/ref=pd_bxgy_b_img_y.

[6] N. Weerasinghe, T. Hewa, M. Dissanayake, M. Yliant-tila, and M. Liyanage, "Blockchain-based Roaming and Offload Service Platform for Local 5G Operators," in *2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC)*, 2021, pp. 1–6. DOI: 10.1109/CCNC49032.2021.9369516.

[7] C. Harris, "Improving Telecom Industry Processes Using Ordered Transactions in Hyperledger Fabric," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6. DOI: 10.1109/GCWkshps45667.2019.9024541.

[8] M. Huillet, *Telefonica, Deutsche Telekom and Vodafone Use Blockchain for Roaming Settlement*, 2020. [Online]. Available: https://cointelegraph.com/news/telefonica-deutsche-telekom-and-vodafone-use-blockchain-for-roaming-settlement (visited on 08/24/2021).

[9] M. A. Islam and M. Jahidul Haque, "Evaluating Document Analysis with kNN Based Approaches in Judicial Offices of Bangladesh," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, 2018, pp. 646–650. DOI: 10.1109/ICCMC.2018.8487847.

[10] M. B.C., A. K., M. Y. M., R. L.R., and S. S.R., "Intelligent Automated Text Processing System - An NLP Based Approach," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 1026–1030. DOI: 10.1109/ICCES48766.2020.9138070.

[11] G. Satyanarayana, J. Bhuvana, and M. Balamurugan, "Sentimental Analysis on voice using AWS Comprehend," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 2020, pp. 1–4. DOI: 10.1109/ICCCI48352.2020.9104105.

[12] A. W. Service, *Amazon Comprehend Developer Guide*, 2021. [Online]. Available: https://aws.amazon.com/comprehend/features/?nc1=h_ls (visited on 07/24/2021).

[13] M. M. Deza and E. Deza, *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2009. DOI: 10.1007/978-3-642-00234-2_1.

[14] J. Santisteban and J. Tejada-Cárcamo, "Unilateral Weighted Jaccard Coefficient for NLP," in *2015 Fourteenth Mexican International Conference on Artificial Intelligence (MICAI)*, 2015, pp. 14–20. DOI: 10.1109/MICAI.2015.9.

[15] S. Gupta, *Overview of Text Similarity Metrics in Python*, 2018. [Online]. Available: https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50 (visited on 07/25/2021).

[16] A. W. Service, *Boto3 - The AWS SDK for Python*, 2021. [Online]. Available: https://boto3.amazonaws.com/v1/documentation/api/latest/index.html (visited on 08/25/2021).

[17] J. X. McKie and R. Liu, *PyMuPDF - the Python bindings for MuPDF*, 2021. [Online]. Available: https://github.com/pymupdf/PyMuPDF (visited on 08/25/2021).

[18] Proximus, *EU Roaming Regulation*, 2021. [Online]. Available: https://www.proximus.be/wholesale/en/id_roaming_eu/public/mobile-roaming/roaming-eu.html (visited on 08/25/2021).

[19] Count-Words-Free, *Compare Text Online*, 2021. [Online]. Available: https://countwordsfree.com/comparetexts (visited on 08/25/2021).

Study To Use Graph-database And Neo4j to Analyze Subscribers' Social Behavior

Nassir Abuhamoud
dept. Electrical and Elecotronic Engineering
Sebha University
Sebha, Libya
nas.abuhamoud@sebhau.edu.ly

Salwa Ali
dept. Network and Communication
Sebha University
Sebha, Libya
sal.ali1@sebhau.edu.ly

***Abstract*— The increasing use of mobile phones plays a major impact on escalating people's interactions more than ever before, Libya is no exception. The scope of this study focuses on understanding users' behavior and studying the possibility to use graph databases to analyze Call Data Records- CDR. We propose a new model to show how graph technologies can be used to analyze CDR to highlight the social behavior of mobile subscribers. This is achieved by using an unsupervised procedure of data analysis and not including the model by any prior knowledge on the applicative context. Data mining outcomes are thus employed for understanding users' behaviors, and present a profile of their characteristics.**

***Keywords*— CDR, Social behavior, neo4j technology, Call.**

I. INTRODUCTION

Population censuses have been widely used in the past for keeping track of the demography and geographical movements of the population [7]. Nowadays, due to short term and everyday mobility, more flexible methods such as registers and indirect databases are employed: CDRs represent an optimal candidate in this sense. One of their main advantages is that they offer a statistically accurate representation of the distribution of people in an area, and they can be used to track large and heterogeneous groups of people [2]. Since CDRs evolve according to the changes of user's behavior, the information they carry "automatically" updates over time. Telecom operators continuously gather a huge quantity of CDRs, from which it is possible to extract additional information with low additional costs, and generate valuable datasets. Analysis of CDR data can successfully be employed in various fields, like monitoring the network, adaptation of supplied services (e.g. customers' billing, network planning) and understanding of the economic level of a certain area. [8].

Neo4j is a Java-based graph database that is open-source. Neo4j's creators describe it as a fully transactional database with a persistent Java engine that allows structures to be stored as graphs rather than tables . Neo4j is widely recognized as the most popular and widely used graph database in the world, with the greatest reference area [1]. The purpose of this research is to examine the characteristics of user behavior using CDR and Neo4j technology. Particularly, Analyzing the social characteristics of a five-month CDR related with Almadar Aljadeed's subscribers. This study's scope is determined by observing subscribers' call activity on a daily basis, subscribers' call activity during the week (working days versus holidays), and overall user behavior over a five-month period. This would aid in identifying peak call times on a daily/weekly basis, as well as where in Wadi Ashati users made more in-calls/text messages and spent longer call duration.

II. BACKGROUND

A. Call Detail Records (CDR)

The cellular (mobile) telephone system is now an established technology. Digital systems have largely superseded older analog systems, and practically all systems now employ the Global System for Mobile-GSM 2G and 3G standards, as well 4G standards. As a result, the infrastructure and operating systems of most mobile phone networks are comparable. Each mobile phone network records and uses information on individual users' phone calls so that they can be billed (or can be debited from a prepaid account). Although the content of a voice or SMS call is not stored, the GSM standard requires that the basic details about the call (also referred to as "metadata") be recorded in the form of a call detail record (CDR). The call and connection made, the time and duration of the call, the use of various services (such as SMS), and the location ID of the cell in use are all included in these records [3]. This description is expanded in the current CDR by include, for example. The connection identifier, the identifier of the caller's base station (BTS), the identifier of the service used, the connection status, and the error code (if any)[6].

Furthermore, the CDR usually has a spatial and temporal resolution, allowing researchers to investigate user behavior by studying CDR data; nevertheless, the CDR is not totally standardized, and will vary depending on the operator, system software provider, and system setup. According to research based on extensive call records, an individual's personality can be deduced from their mobile phone usage pattern [13]. Figure 1 shows the construction of CDR.

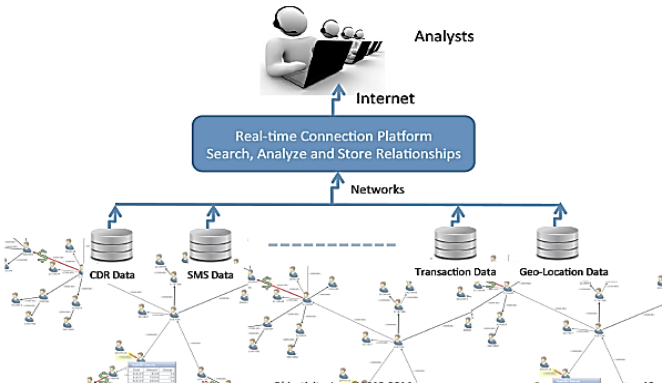


Fig. 1. CDR Construction

B. Graph Database

Graph databases are databases that store data structures that are graph orientated [12]. A graph is a representation of a set of items in which links connect some pairs of objects. Each node represents an entity (a person, place, thing,

category, or other piece of data), and each relationship describes how two nodes are connected [15]. A graph is a representation of a set of objects where some pairs of objects are connected by links. It is composed of two elements: a node and a relationship, where each node represents an entity (a person, place, thing, category or other piece of data), and each relationship represents how two nodes are associated[10].

1) Neo4j Technology

Neo4j is a Java and Scala-based open-source NoSQL graph database that provides Atomicity, Consistency, Isolation, and Durability (ACID). Neo4j technology has a number of advantages, including the ability to scale to billions of nodes and relationships, as well as a simple user interface and copious free internet resources. In addition, Neo4j has a large technical support community[11]. Furthermore, when large amounts of interconnected data must be drawn, Neo4j is the ideal option. The reason being Neo4j is much faster than traditional databases and it analyzes and traverses all data in real time. Additionally, Neo4j provides an effective graph experience by traversing millions of nodes in milliseconds [12]. For technical support, Neo4j has a large community. Furthermore, when large amounts of interconnected data must be drawn, Neo4j is the ideal option. Because Neo4j is faster than traditional databases and analyzes and traverses all data in real time, it is a good choice. Furthermore, Neo4j provides an effective graph experience by traversing millions of nodes in milliseconds [6].

2) Cypher Query:

Cypher is an expressive (yet compact) graph database query language. Cypher is designed to be easily read and understood by developers, database professionals, and business stakeholders. Its ease of use derives from the fact that it is in accord with the way we intuitively describe graphs using diagrams [11]. Cypher enables a user (or an application acting on behalf of a user) to ask the database to find data that matches a specific pattern [10].

C. Wadi Ashatii Distric:

Wadi al Shati, sometimes referred to as Ashati, is one of the districts of Libya in the central-west part of the country. The Depression extends eastward from the city of Ashkeda to the city of Ideary in the west, a distance of about 140 kilometers in the west. According to estimates, by 2020, the total population of the region was 157,747 with 150,353 Libyans. Figure 2 shows the location of Wadi Ashatii district in Libya Map.

Fig. 2. Wadi Ashatii District

D. Social behavior

Mobile phones are becoming more and more ubiquitous and powerful. Its ability to record information about call activity provides a unique opportunity to specify how such information could be used to shape people's call response behavior in call logs (e.g. specific time of call, date of call, type of call, duration of call) [5]. In terms of social behavior, Social behavior of mobile phone users describes social interactions and activity patterns. Social interactions, which mostly take two forms: temporal and spatial, involve features that represent user's interactions and their activities with other individuals. Call log data can show temporal interaction whereas Bluetooth data can reveal spatial interaction. Activity patterns; however, are related to the spatial area of the daily activities of users such as GPS , WLAN data [9].

III. METHODOLOGY

This study is conducted to investigate the social behavior of mobile phone users in the Wadi Ashati district based on CDR and Neo4j technology. Relevant analysis on subscribers' activity is performed so that demographic characteristics could be identified. The Almadar Aljadeed based CDR undergoes a preparation process to represent the subscribers' data during five months . This is followed by a coding stage using Python. Subsequently, users' characteristics could easily be acknowledged. The data would then be processed in the Neo4j Program ,where suitable queries are applied using Cipher queries. The following figure (Fig. 3) depicts the graph model that represents a CDR record. It illustrates how the CDR features are represented in the graph.

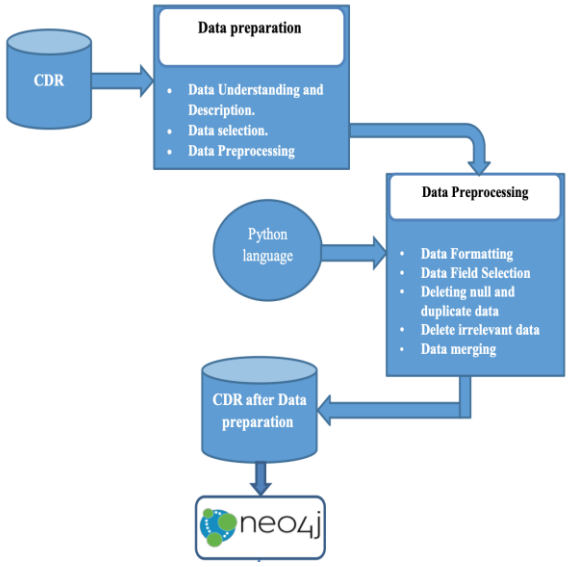


Fig. 3. Architecture of model

A. CDR

The CDR describe the metadata of each telecommunication event including the phone numbers of a caller and receiver, the start date and time of an event and the duration. Current CDRs expand this description by adding for e.g. the identifier of a connection, the identifier of a Base Transceiver Station (BTS) used by the caller, an identifier of a service used [6]. The data related to mobile phone activity of AMADAR operator was collected over a period of five months in 2017, where the city of Wadi Ashati was selected as a desired area for investigation. As can be seen from Figure 4, the area of analysis is divided into three different subareas, named A, B and C.

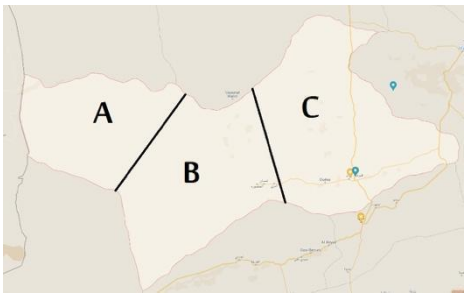


Fig. 4. Regions of Wadi Ashatii

We investigated the CDR based on aggregated features within the predefined areas over a five month period . such features includes: the overall in-calls, text messages and call duration . The following table reveals the number of BTS in subareas.

Table 1: number of BTS per Area

Area	BTS.no
A	12
B	5
C	7

B. Data Preparation

The CDR data used in this investigation belongs to Almadar Aljadid operator, where the records were collected over a period of five months(from February 2017 to June 2017). Around 52 million records were found in every daily file, and this results in difficulties in conducting data preparation. In order to deal with this problem, we divide the file to 152 separate sub files using the Python language, each of which contains million records. This could help in implementing the first stage with all the required steps to perform data preparation.

C. Data Understanding

This phase is necessary in order to extract relevant data for the analysis. It entails a number of tasks, beginning with the selection of appropriate fields from the CDR, followed by an examination of the contained values and an evaluation of their significance for the study. In this study , the CDR records had 73 different fields, some of which were discovered to be null; others, however, had duplicated data, such as the caller's number or the recipient's number, and occasionally the duration of the calls..

D. Data Selection

The data field selection process is crucial, and requires great knowledge and experience to select useful and robust features. The fields have been chosen based on studying normal behavior of users including :caller numbers; recipient numbers; caller location; call duration and calls' type(voice call or an SMS message), Figure 5 reveals the result of execution of the code. Only useful data is produced including call time, calling number, called number, location of calling number, location of called number, call duration, exchange and type of services.

E. Data Preprocessing

This step requires more attention to set the data in a suitable form to be processed into the Neo4j software, and perform relevant queries. It consists of: formatting data; selecting suitable fields for the study; deleting spaces and duplication; eliminating irrelevant data within the selected fields and lastly putting data in a final form. Figure 5 indicates the data preprocessing steps.

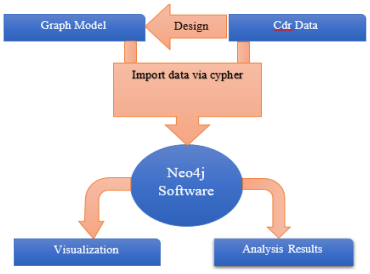


Fig. 5. Data Preprocessing [12]

As the CDR records are imported into a histogram database, we have created a number of queries to discover the abnormal behavior. the behavior is considered as abnormal if the user's behavior was significantly different from their regular behavior .The following are scenarios that could be considered for social behavior in the predefined sub-areas:

1) In-call based behavior

In this scenario, we will define the location of users who make more in-calls .Hence, the result highlights a place where users have a broad social network and strong relationship with others.

2) SMS based behavior

This scenario highlights a location where educated people use more text messages, while less educated users reside in locations with low SMS rate.

3) Duration call based behavior

In this scenario, we will check whereabouts users make long call duration in order to determine the location of more talkative subscribers.

IV. RUNNING MODEL

In order to study the behavior of subscribers, it is crucial to build a model for the data prior to being imported. There are various methods to import data in Neo4j, but the most common way is by uploading the data as a csv file. Load CSV operator, which is used for small/medium size datasets(up to 10 million records) is built into Neo4j.The data import process consists of two main steps; the first step is identifying the graph pattern representing the phone call. This step is critical because identifying unique phone numbers and their relationships in tabular data is hard and sophisticated. To perform the first step, we utilize phone calls data to build a graph where the phone numbers are connected by phone calls. Figure 6 elucidates the graph pattern of phone calls. It is apparent that a single phone call connects together four entities: 2 phone owners, a location (the cell site, the caller was next to when he/she initiated the call), a city and its surrounding regions.

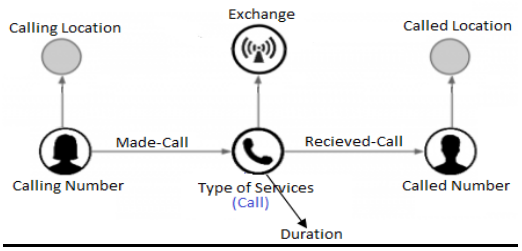


Fig. 6. Graph model to represent the phone calls.

As the data related to subscribers' identity is not shown in the CDR, the model is built based on the spreadsheet data . To store our graph, we will use Neo4j, a popular graph database. The next step to build our model is Using query language (cypher query) to fill in CDR data in neo4j. As mentioned earlier, Neo4j has a language called Cypher that makes it easy to import csv files. Figure 7 demonstrates a sample of transforming the CDR data into a Neo4j graph. The code transforms the original table into a new table using Neo4j.

```
CREATE (CallinNode)-[:Made_Call]->(CALL)-[:received_call]->(CalledNode),
(f)-[:FRM_BTS]->(CallinNode), (g)-[:TO_BTS]->(CalledNode).
CREATE (CallinNode)-[:Made_Call]->(CALL)-[:received_call]->(CalledNode),
(Exchange)-[:FROM]->(CALL), (s)-[:DURATION]->(CALL).
```

Fig .7. A sample code of CDR transformation

V. RESULT AND DISCUSSION

This section describes the outcome of our analysis on the CDR related to subscribers in the city of Wadi Ashati. It is assured that the proportion of Almadar Aljadeed-based SIM making calls/text messages in Wadi Ashati during the investigation period was approximately 6% of the overall officially recorded Almadar Aljadeed- based SIM in Libya . Surprisingly, the latest records of Libyans population showed that 1.8 % of total populations reside in Wadi Ashati. Therefore, it is estimated that the number of SIM cards exceeds the actual number of Libyan subscribers and thus more emigrants reside in Wadi Ashati. As we divided the city to three different sub-areas, the analysis could determine which sub-area is more likely to make more in-calls, text message making calls/SMS .

Figure 8 demonstrates the rate of in-calls within the predefined sub-areas. Our result showed that the Almadar Aljadeed subscribers within the city of Wadi Ashati made about 6% of overall in-calls in Libya. It is worth noting that users in sub-area B are more likely to make in-calls- approximately 45% of total calls in Wadi Ashati, was made by users in region B. This reflects B as an actively central area in the city where the social network is wide enough for subscribers to reach more contacts.

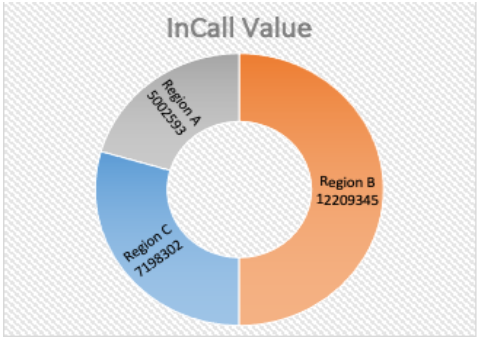


Fig.8. In-calls rate in sub- areas

Figure 9 illustrates the rate of text messages made by subscribers in the sub-areas A, B and C. Only 5% of total SMS had been sent by users in sub-area B. The lowest rate of text-messages in sub-area B highlights the level of educated people in this sub-area compared to other locations. When it comes to subscribers' characteristics, this group has a preference for phone-calls over text-messages. Considering the previous figure, it is noteworthy that higher rate of in-calls in sub-area B might result in the network being busy most of the time in which people would send SMS as an alternative way to reach their contacts. The rate of text messages for both sub-area A and C shows relatively similar behavior. As can be noticed from the figure, sub-area A and C are also considered the most places in Wadi Ashati where subscribers have a tendency to text-messages more than other users (about 47% and 48% of total text-messages in Wadi Ashati were sent from sub-area A and C respectively). This undeniably confirms that more educated people in Wadi Ashati are resided in these locations

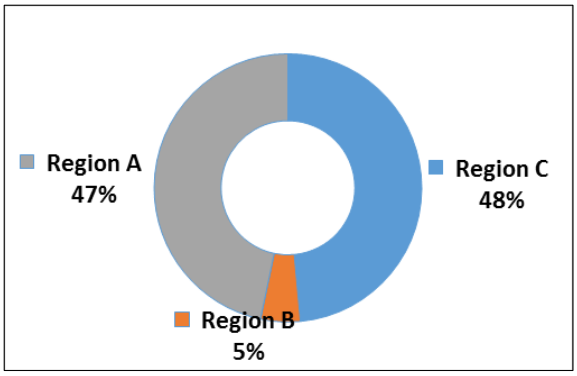


Fig.9. Rate SMS in sub-areas

Figure 10 reveals the rate of total local in-calls made by users in the sub-area A, B and C. half of total in-calls within the sub-area C were made locally inside the district. Additionally, our analysis highlights C the most sub-area in Wadi Ashat that makes in-calls to the main BTS in Tripoli. This also confirms C the most place in Wadi Ashat where residents have strong tie relationships. This finding could be an indicator to consider this location to have more talkative subscribers. Such results would definitely help decision makers to target relevant users for special offers; thereby, rising their profits. Nevertheless, the rate of overall local in-calls in both sub-area A and B was recorded at 43% and 62% respectively.

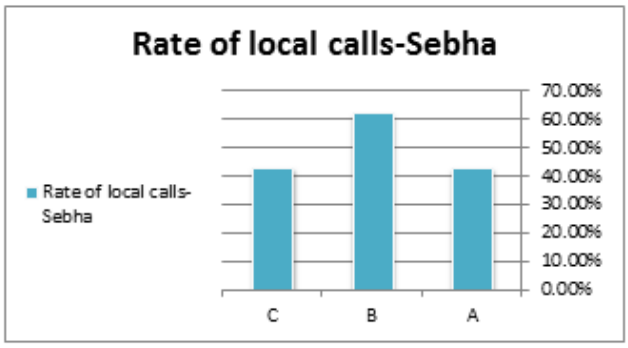


Fig. 10. Rate of local calls in sub-areas

With regard to the duration of calls, the sub-area A presents the highest rate of average duration call-reaching 23 minutes and 21 seconds. Considering our previous results, the Sub-area A recorded the highest rate of call duration despites the fact that the percentage of subscribers' local call was low. This means that residents had higher social relationships with people outside the city.

What is more, it was found that mobile phone activities varied according to usage time. Based on our analysis of the CDR on working days, the rush time of mobile activity was recorded at (12pm- 3pm). Following this, the midnight time between 11pm and 1am. This outcome indicates that the city is less active than the average of normal levels

VI. CONCLUSION AND FUTURE WORK

This study examined the CDR related to the Almadar Aljadeed subscribers over a period of five months. The CDR underwent a preparation process in order to highlight mobile phone activities within Wadi Ashati District. Relevant queries were implemented using Cipher queries with the neo4j software. The findings were highlighted based on the analysis of subscribers' activity in three different sub-area within Wadi Ashati. The mobile phone features of interest included rush time, in-call rate, calls and text messages rate. Our result defined the location of high-level educated users based on text messages activity. The outcome also confirmed the correlation between call duration and subscribers' characteristics. This would assist the network operators to target a specific group of users for special offers. Future works could analyze social characteristics of subscribers in urban and rural areas. Further studies are also recommended to examine the CDR of different cities and compare their social behavior with total population.

REFERENCES

[1] Barbosa H et al. "Human mobility: models and applications". Phys. Rep. 2018 . 734, 1–74.

[2] M. Zignani, C. Quadri, S. Gaito, GP Rossi. "Calling, texting, and moving: multidimensional interactions of mobile phone users". *Computational Social Networks* 2, 2015, no. 1, 1-24

[3] D. Spörle and M .Kajo, "Surveying the depth of user behavior profiling in mobile networks". *Network*, 2019, *17*

[4] K. Dasgupta, R .Singh, B.Viswanathan, D .Chakrabort,S. Mukherjea, , A.A .Nanavati,. and A.Joshi,"Social ties and their relevance to churn in mobile telecom networks". In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, March 2008 (pp. 668-677).

[5] Wang, Feilong, and Cynthia Chen. "On data processing required to derive mobility patterns from passively-generated mobile phone data." *Transportation Research Part C: Emerging Technologies* 87, 201,: 58-74.

[6] Eugenia.Politou, Efthimios Alepis, and Constantinos Patsakis."A survey on mobile affective computing". *Computer Science Review* 25 ,2017: 79-100.

[7] Lasantha.Fernando, Aparna Surendra, Sriganesh Lokanathan, and Thavisha Gomez, Predicting population-level socio-economic characteristics using Call Detail Records (CDRs) in Sri Lanka." In *Proceedings of the Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*,2018, pp. 1-12.

[8] Zhiren.Huang, Ximan Ling, Pu Wang, Fan Zhang, Yingping Mao, Tao Lin, and Fei-Yue Wang, "Modeling real-time human mobility based on mobile phone and transportation data fusion", *Transportation research part C: emerging technologies* 96 ,2018, 251-269.

[9] Phithakkitnukoon, S. and Smoreda, Z, "Influence of social relations on human mobility and sociality: a study of social ties in a cellular network. *Social Network Analysis and Mining*, *6*(1), 2016, pp.1-9

[10] Nassir.Abuhamoud and Emsaieb Geepalla,"Analysis CDR for Crime Investigation using graph-based method (Neo4j)." In *International Conference on Technical Sciences (ICST2019)*,2019, vol. 4, p. 06. 2019.

[11] Han.Wang, and Liam Kilmartin.2014,"Comparing rural and urban social and economic behavior in Uganda: insights from mobile voice service usage." *Journal of Urban Technology* 21, 2 (2014): 61-89.

[12] Nassir.Abuhamoud, Ibrahim Alsadi, and Salwa Ali,"Detecting SIMBox Fraud Using CDR Files And Neo4j Technology." In *2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, 2021,pp. 259-263.

[13] Wang, Xu, Zimu Zhou, Fu Xiao, Kai Xing, Zheng Yang, Yunhao Liu, and Chunyi Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis." *IEEE Transactions on Mobile Computing* 18, 2018, no. 9, 2190-2202

[14] Rodrigues, André Leite, Mariana Giannotti, Matheus HC Cunha Barboza, and Bianca Bianchi Alves. "Measuring mobility inequalities of favela residents based on mobile phone data.", *Habitat International* 110, 2021,102346.

[15] Drobyshevskiy, Mikhail, and Denis Turdakov. "Random graph modeling: A survey of the concepts", *ACM Computing Surveys (CSUR)* 52, 2019, no. 6 , 1-36

Diagnosing Pneumonia using Convolutional Neural Networks

Walid Abdalla Ramdhan Abdalla
Department of Computer Science
College of Electrical and Electronics Technology-Benghazi
Benghazi, Libya
walid.tajuri@gmail.com

***Abstract*—Pneumonia is a contagious lung illness brought on by tiny organisms. Pneumonia is caused by transmissible microorganisms, which means it may spread from person to person. It is still the leading infectious illness killer of children under the age of five. Microbiology testing may be dishonestly negative or, as is usually the case with patients who are continually ventilated, may be mistakenly certain due to continuous colonization of the respiratory system with bacteria. In recent years, artificial intelligence and deep learning have become increasingly popular in medicine. The use of deep learning in the analysis of medical images improves consistency and increases reporting accuracy. The study presented in this paper uses a convolutional neural network to construct a processing model to aid in the classification challenge of determining whether a chest X-ray exhibits alterations associated with pneumonia or not. As the model trains, the accuracy improves and the loss reduces. Dropout regularization and data augmentation are used to prevent overfitting. Thus, the suggested deep learning models produce efficient and convincing findings.**

***Index Terms*—Artificial Intelligence, Convolutional Neural Networks, Deep Learning**

I. INTRODUCTION

Pneumonia is a disease that can strike at any time in a person's life. Pneumonia is caused by microorganisms and viruses. it affects nearly 200 million people worldwide, with 100 million adults and another 100 million children under the age of five. Each year, approximately 4 million people die prematurely as a result of air pollution diseases such as pneumonia, according to World Health Organization statistics. Pneumonia has a greater impact in low-income countries, where it is the leading cause of mortality and when access to diagnostic and therapeutic offices is limited. Because of several factors including imprecise X-ray images or confusion with different disorders, the diagnosis may not always be particularly successful. On the other hand, analyzing X-ray images can be tedious, time-consuming, and difficult without access to specialized expertise, which may not be available in smaller towns. As a result, computer-aided diagnostics (CAD) enters the images. CAD has emerged as the most credible method for detecting breast cancer, lung cancer, and pneumonia [4].

According to current studies, deep learning, particularly, convolutional neural network (CNN), has emerged as the most acclaimed technique for image classification [1,2,10]. Deep learning has the ability to reshape illness diagnosis by creating a classification model that can identify images, which is a difficult challenge even for professional radiologists [3].

This research presents a methodology for the classification of pneumonia based on chest X-ray images. Several chest X-ray images were classified as normal or pneumonia by a few artificial CNNs. The dataset must be divided into three phases for training: train, test, and validation. The training dataset includes 5856 chest X-ray images, the test dataset includes 624 images, and the validation dataset includes 16 images. The binary classifiers are used in our research. The binary classifier is divided into two groups: 0 (the default) and 1(abnormal). Patients with pneumonia fall into the category of class 1, whereas those who are normal fall into the category of class 0. We want to build a deep learning classifier model that can categorize the images we give it as input images, and once trained, the classifier will be able to distinguish between those images. Finally, we will validate and test the model. Our model has an accuracy of 92%.

II. RELATED WORK

For many years, the identification of pneumonia using chest X-rays has been an open subject [15, 16], with the primary constraint being a lack of publicly available data. Traditional machine learning approaches have been widely researched. Chandra et al. [17] derived eight statistical variables from lung areas segmented from chest X-ray images and utilized them to categorize them. They used five standard classifiers: MLP, random forest, sequential minimum optimization (SMO), classification by regression, and logistic regression. They tested their approach on 412 photos and attained an accuracy rate of 95.39% using the MLP classifier. Kuo et al. [18] investigated 11 characteristics to diagnose pneumonia in 185 individuals with schizophrenia. They used these characteristics in a variety of regression and classification models, including decision trees, support vector machines, and logistic regression, and then compared the results. Using a decision tree classifier, they attained the greatest accuracy rate, 94.5%; the other models fell short by huge percentages. Yue et al. [19] employed six characteristics to diagnose pneumonia in chest CT scan images of 52 patients, with a best AUC value of 97. These strategies, however, are not generalizable and were tested on tiny datasets. Unlike machine learning algorithms, which require handcrafted features to be extracted and selected for

classification or segmentation [20,21], deep learning-based methods perform end-to-end classification [22,23], in which relevant and informative features are automatically extracted and classified from input data. CNNs are popular for image data categorization because they automatically extract translationally invariant features from the input images and filters. CNNs are transcriptionally invariant and outperform machine learning and classical image processing approaches in image classification tasks, making them popular among academics. Sharma et al. [24] and Stephen et al. [25] developed basic CNN architectures for pneumonic chest X-ray image categorization. To compensate for the scarcity of data, they employed data augmentation. On the dataset supplied by Kermany et al. [26] Sharma et al. acquired a 90.68% accuracy rate while Stephen et al. obtained a 93.73% accuracy rate. However, data augmentation only supplies a limited quantity of additional information from which CNNs may learn, and so may not considerably improve their performance. Rajpukar et al. [27] employed the DenseNet-121 CNN model for pneumonia classification but only got a 76.8% f1-score. They thought that the lack of patient history was a key contributor to the poor performance of both their deep learning model and the radiologists with whom they compared the method's performance. Janizek et al. [28] suggested an adversarial optimization-based method for removing model reliance on dataset source and producing robust predictions. They got a 74.7% AUC in the source domain and a 73.9% AUC in the target domain. Zhang et al. [29] suggested a confidence-aware module for detecting abnormalities in lung X-ray images, treating the detection job as a one-class issue (determining only the anomalies). On their dataset, they obtained an AUC of 83.61%. Tuncer et al. [30] employed a machine learning-based strategy in which they performed the fuzzy tree modification to the photos before dividing them into exemplars. The features were then retrieved using a multikernel local binary pattern, and the data were classed using standard classifiers. They tested the approach on a small dataset of COVID-19 and pneumonia samples and discovered that it had a 97.01% accuracy rate. Transfer learning, in which information learned from a big dataset is applied to fine-tune the model on a current small dataset, is now a popular strategy for addressing the data scarcity issue in biomedical image classification problems. Recently, Wang et al. [31], Chaudhary and Pachori [32], Ibrahim et al. [33], and Zubair et al. [34] employed solely transfer learning techniques to classify pneumonia using various CNN models pre-trained on ImageNet [35] data. The majority of cutting-edge deep learning algorithms for pneumonia diagnosis rely on a single CNN model. Ensemble learning [36,37] fuses the judgments made by many CNN models, effectively combining the prominent characteristics of all its base models in the ensemble model, gathering complementing information from the distinct classifiers, and providing a more robust conclusion. This paradigm has received little attention in respect to the pneumonia detection challenge. Jaiswal et al. [38] employed a mask region-based CNN for pneumonia trace identification through segmentation, using a detector RetinaNet and SeResNext101 encoders. For images thresholding, an ensemble model comprised of ResNet-50 and ResNet-101 was used. Gabruseva et al. [39] suggested a deep learning architecture based on a single-shot for the localisation of pulmonary opacity. During the training phase, they performed an ensemble of several checkpoints (snapshot ensembling) and achieved a mean average precision (mAP) of 0.26 over several intersection over union thresholds, one of the best results in the Radiological Society of North America (RSNA) Pneumonia Detection Challenge. Pan et al. [40] employed an ensemble of the Inception-ResNet v2, XceptionNet, and DenseNet-169 models for pneumonia diagnosis and attained the best result in the challenge, a mAP value of 0.33, on the same test. To the best of our knowledge, ensemble models have not been utilized for classification tasks in the pneumonia detection issue, and in this work, we employed ensemble learning for the first time in this domain to classify lung X-rays into "Pneumonia" and "Normal" classes. Three cutting-edge CNN models with transfer learning, GoogLeNet, ResNet-18, and DenseNet-121, were utilized to construct the ensemble utilizing a weighted average probability strategy with a unique approach to weight allocation. This study ameliorates the problem of poor image resolution and enhances model performance by avoiding the well-known over-fitting effect on model performance.

III. MATERIALS AND METHODS

A. The Dataset

The dataset used for this study was provided by the Guangzhou Centre for Obstetrics and Gynecology[5]. The data set includes 5856 chest X-rays. It is divided into three parts: Train, Val, and Test, which are used as training, validation, and test data. In this study, 80% of the images are used as training data, 10% as validation data, and 10% as test data. Each of these three sections contains two sub-folders with images diagnosed as pneumonia or normal. The number of images of both categories (normal and pneumonia) in the train, test and Val dataset are:

TABLE I
THE NUMBER OF IMAGES BELONGING TO BOTH CLASSES

Dataset Type	Normal	Pneumonia
Training	1341	3875
Test	234	390
Validation	8	8

These images are of high quality and vary in size, but were then resized to train the model. Figure 1 shows several examples of X-ray images and their corresponding labels.

B. Dataset Preparation

Usually, the first step in building a model is to pre-process the data we input. The original images are RGB, but in this experiment, they are scaled to a size of 200 x 200 pixels. Then normalize the pixel density value by dividing the pixel value by 255. In this way, the pixels in the image are represented by floating-point numbers between 0 and 1, rather than integers

Fig. 1. Several examples of X-ray images and their corresponding.

between 0 and 255. This should have a positive impact on CNN performance [6].

C. Convolutional Neural Network and Tools Selection

In general, artificial intelligence is an information processing paradigm that attempts to replicate the human brain. CNNs represents an enormous breakthrough in image recognition and classification, where they're most ordinarily used. They contain few layers, like input, output and between them, there are hidden layers. These hidden layers do the foremost add terms of calculation. Convolutional layers are found within the hidden layers [4], [7].

Hidden layers also can accommodate pooling layers and fully connected layers. The convolutional layer is the most critical building element of a CNN [9]. This is often referred to as a feature detector or a feature extractor. Convolutional scans a given image using a kernel matrix and applies filters to get a certain effect. In artificial intelligence, parts are used to include extraction to choose the most significant pixels in an image. Each neuron in the convolutional layer is only connected to a small number of neurons in the convolutional layer that follows [7]. This type of architecture allows the network structure to concentrate on a tiny low-level characteristic in the first hidden layer, then aggregate it into higher-level features in the next hidden layer, and so on [7]. Pooling layers are used after the convolutional layer to decrease the size of the input images without losing any critical information. This is done to decrease the computational cost as well as the amount of memory used. Pooling causes the model to sum up by avoiding overfitting and improves computing efficiency while preserving the highlights. In our studies, we utilized the ReLU activation function because the feature map's non-linearity necessitates the employment of ReLU layers It also makes feature maps sparser, and performs well in deep neural networks, owing to the fact that it does not saturate for positive values and is quick to calculate. If the values in the feature map are less than zero, they are set to zero; however, if the values are larger than zero, the same value is passed. In addition, a dropout approach was utilized in the model construction to improve performance. Dropout is a method in which certain neurons are shut off at random and are not used for that iteration. The building blocks of a classical CNN model is illustrated in Figures 2.

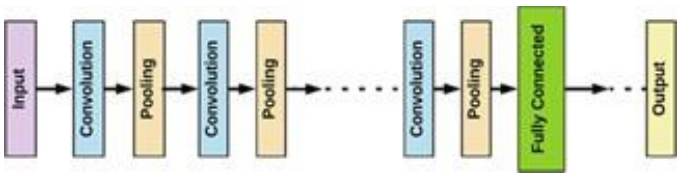


Fig. 2. Building blocks of a classical CNN model(Sultana et al., 2018).

IV. ENVIRONMENT IMPLEMENTATION

The CNN model was done with Python programming and libraries such (Sklearn, Matplotlib, Keras, TensorFlow, and NumPy) [15]. The training and testing of the model were performed on google colab.

V. RESULT ANALYSIS

The CNN model for identifying pneumonia patients using X-ray images is presented in this work. To train and evaluate the models, a higher number of X-ray images were gathered. Dropout was employed for regularization, and a balance between too much and not enough network capacity was obtained. The influence of critical hyperparameters such as batch size, number of convolutional layers, activation function, image resolution, the size of each layer, number of epochs, learning rate, and neurons in the dense layer were investigated using recent deep learning approaches.

Transfer learning approaches such as those used by Wang et al. (2020) achieved 93,3 percent accuracy using a pre-trained model based on ImageNet and 11,75 million parameters, while Chaudhary and Pachori (2020) achieved 98,3 percent accuracy using 750 total images but without specifying the network capacity, trainable parameters, or computational complexity for the final model. Because of the large network capacity (million parameters) compared to the tiny number of X-ray images, the generated results are biased and overfitted in all of these circumstances. The pre-trained models had little resemblance to X-ray images and required extensive fine-tuning to make them relevant to the situation at hand, however this resulted in models with excess capacity relative to sample size. Another disadvantage of these technologies is that they result in highly large and heavy models, making them difficult to deploy on devices with limited capacity. Other non-transfer learning approaches, such as Tuncer et al. (2021), achieved 98,08 percent accuracy with 17 convolutional layers and 164 million parameters. The model's network capacity or trainable parameters are not mentioned.

Most earlier studies did not execute the hyperparameter optimization procedure in a thorough and automated manner, as this study does. In this research, we got the best outcome when trained for 60 epochs, the trained model produced outstanding results. The accuracy score was 0.9230, and the training loss was 0.2317 as a consequence. The training loss is greatly reduced as a result of data augmentation and dropout regularization, which allows the model to learn better with greater variance of images rather than relying on weights to remember the data. Furthermore,the network capacity was

adjusted to only 3,453,634 parameters. Figure 3 shows the summary impression of the CNN model used for the experiments described in this study.

```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 148, 148, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 74, 74, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 72, 72, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 36, 36, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 34, 34, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 17, 17, 128)       0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 15, 15, 128)       147584
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 7, 7, 128)         0
_________________________________________________________________
dropout (Dropout)            (None, 7, 7, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 6272)              0
_________________________________________________________________
dense (Dense)                (None, 512)               3211776
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 1026
=================================================================
Total params: 3,453,634
Trainable params: 3,453,634
Non-trainable params: 0
```

Fig. 3. Model Summary.

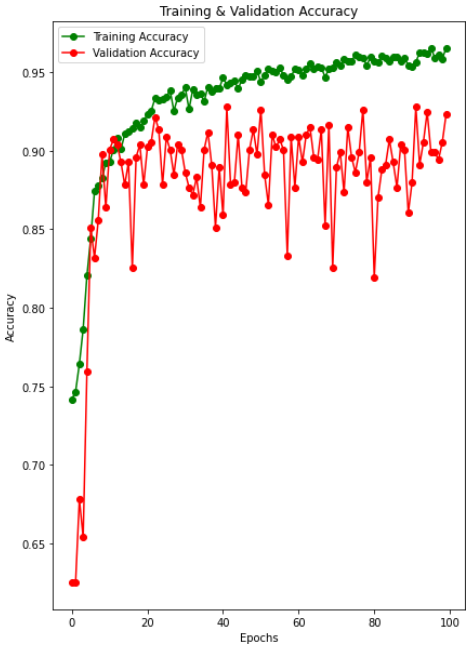


Fig. 4. Training and validation accuracy.

VI. CONCLUSION

In medicine, deep learning plays an important role in diagnosing diseases such as lung cancer, pneumonia, and breast cancer. The aim of our project is to diagnose chest X-ray pneumonia using CNNs. The proposed CNN model was

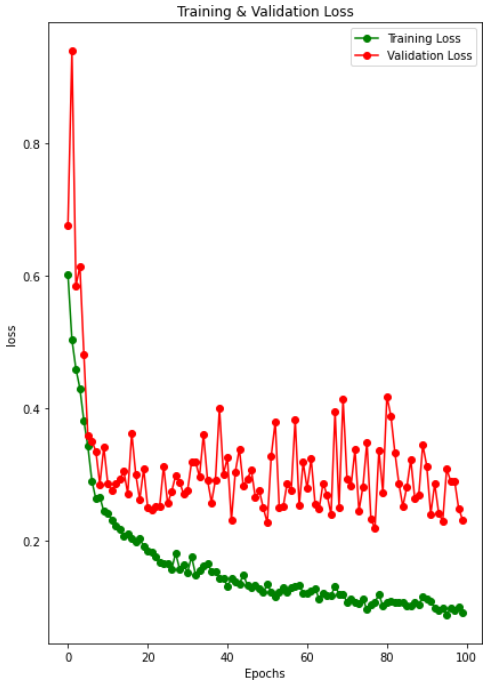


Fig. 5. Training and validation loss.

```
In [39]: result = model.evaluate_generator(test_data_gen, steps=len(test_data_gen), verbose=1)
         print('Loss:', result[0])
         print('Accuracy:', result[1])

7/7 [==============================] - 4s 566ms/step - loss: 0.2318 - accuracy: 0.9231
Loss: 0.23177196085453033
Accuracy: 0.9230769276618958
```

Fig. 6. Printing Accuracy.

created to deliver accurate diagnoses, and its accuracy has improved over previous studies. The proposed model avoided overfitting by (1) expanding the size of the data set (through data augmentation methods), (2) including regularization techniques such as dropout, and (3) conducting hyperparameter tuning. The effect of key hyperparameters was investigated. Previous studies have shown that the produced findings are biased and overfitted as a consequence of the large network capacity compared to the limited number of X-ray images, and hyperparameter tuning is done weakly and manually in just a few hyperparameters.

The proposed model achieved a respectable accuracy of 92.30 percent for binary classification, outpacing previous research in the literature. Finally, the proposed models lowered network capacity and size to the greatest extent feasible, allowing the final models to be deployed locally on devices with restricted capabilities.

In the future, the proposed model could be extended to detect the presence of COVID-19 in patients' frontal X-ray images. Identifying frontal X-ray images containing Coronavirus and pneumonia has become a major difficulty in recent years, and our next strategy will address this issue.

REFERENCES

[1] A. Esteva, "A guide to deep learning in healthcare"," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.

[2] D. Ravi, "Deep Learning for Health Informatics," IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 4–21, 2017.

[3] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology"," Nature Reviews. Cancer, vol. 18, no. 8, pp. 500–510, 2018.

[4] L. L. G. Oliveira, S. A. Silva, L. H. V. Ribeiro, R. M. D. Oliveira, C. J. Coelho, and A. L. S. Andrade, "Computer-aided diagnosis in chest radiography for detection of childhood pneumonia," Int. J. Med. Inf, vol. 77, pp. 555–564, 2008.

[5] D. Kermany, K. Zhang, M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, v2, 2018. Accessed July 2020. [Online]. Available: http://dx.doi.org/10.17632/rscbjbr9sj.2.

[6] R.C. Gonzalez, R.E. Woods. "Digital Image Processing", Prentice Hall, 2007, pp 85.

[7] A. Geron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", O'Reilly Media, Inc., Canada, 2019.

[8] Chauhan, P., Sharma, N., Sikka, G., The emergence of social media data and sentiment analysis in election prediction. J. Ambient Intell. Hum. Comput., 1,1–27, 2020.

[9] Sharma, S., Juneja, A., & Sharma, N. Using deep convolutional neural network in computer vision for real-world scene classification. In 2018 IEEE 8th International Advance Computing Conference (IACC), 1, 284-289, IEEE, December, 2018

[10] Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., Mittal, A., Pneumonia detection using CNN based feature extraction, in: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, February, IEEE, pp. 1–7.

[11] Prayogo, K. A., Suryadibraya, A., & Young, J. C. Classification of pneumonia from X-ray images using siamese convolutional network. Telkomnika, 18, 3, 1302–1309, 2020.

[12] Gabruseva, T., Poplavskiy, D., Kalinin, A., Deep learning for automatic pneumonia detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 350–351, 2020.

[13] Gabruseva, T., Poplavskiy, D., Kalinin, A., Deep learning for automatic pneumonia detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 350–351, 2020.

[14] Janizek, J.D., Erion, G., DeGrave, A.J., Lee, S., II, An adversarial approach for the robust classification of pneumonia from chest radiographs, in: Proceedings of the ACM Conference on Health, Inference, and Learning, 2020, April, pp. 69–79.

[15] Albahli S., Rauf H., Algosaibi A. & Balas V. AI-driven deep CNN approach for multi-label pathology classification using chest X-Rays. PeerJ Computer Science. 7 pp. e495 (2021) https://doi.org/10.7717/peerj-cs.495 PMID: 33977135.

[16] Albahli S., Rauf H., Arif M., Nafis M. & Algosaibi A. Identification of thoracic diseases by exploiting deep neural networks. Neural Networks. 5 pp. 6 (2021).

[17] Chandra T. & Verma K. Pneumonia detection on chest X-Ray using machine learning paradigm. Proceedings Of 3rd International Conference on Computer Vision and Image Processing. pp. 21-33 (2020).

[18] Kuo K., Talley P., Huang C. & Cheng L. Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. BMC Medical Informatics and Decision Making. 19, 1–8 (2019) https://doi.org/10.1186/s12911-019-0792-1 PMID: 30866913.

[19] Yue H., Yu Q., Liu C., Huang Y., Jiang Z., Shao C., et al. & Others Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. Annals Of Translational Medicine. 8 (2020) https://doi.org/10.21037/atm-20-3026 PMID: 32793703.

[20] Meraj T., Hassan A., Zahoor S., Rauf H., Lali M., Ali L., et al. Lungs nodule detection using semantic segmentation and classification with optimal features. Preprints. (2019).

[21] Rajinikanth V., Kadry S., Damaseˇvičius R., Taniar D. & Rauf H. Machine-Learning-Scheme to Detect Choroidal-Neovascularization in Retinal OCT Image. 2021 Seventh International Conference on Bio Signals, Images, And Instrumentation (ICBSII). pp. 1-5 (2021).

[22] Kadry S., Nam Y., Rauf H., Rajinikanth V. & Lawal I. Automated Detection of Brain Abnormality using Deep-Learning-Scheme: A Study. 2021 Seventh International Conference on Bio Signals, Images, And Instrumentation (ICBSII). pp. 1-5 (2021).

[23] Rajinikanth V., Kadry S., Taniar D., Damaseˇvičius, R. & Rauf H. Breast-Cancer Detection using Thermal Images with Marine-Predators-Algorithm Selected Features. 2021 Seventh International Conference on Bio Signals, Images, And Instrumentation (ICBSII). pp. 1-6 (2021).

[24] Sharma H., Jain J., Bansal P. & Gupta S. Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). pp. 227-231 (2020).

[25] Stephen O., Sain M., Maduh U. & Jeong D. An efficient deep learning approach to pneumonia classification in healthcare. Journal Of Healthcare Engineering.2019(2019) https://doi.org/10.1155/2019/4180949 PMID: 31049186

[26] Kermany D., Zhang K. & Goldbaum M. Labeled Optical Coherence Tomography (OCT) and Chest Xray Images for Classification. (Mendeley,2018).

[27] Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., Duan T., et al. & Others Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. ArXiv Preprint ArXiv:1711.05225. (2017).

[28] Janizek J., Erion G., DeGrave A. & Lee S. An adversarial approach for the robust classification of pneumonia from chest radiographs. Proceedings Of the ACM Conference on Health, Inference, And Learning. pp. 69-79 (2020).

[29] Zhang J., Xie Y., Pang G., Liao Z., Verjans J., Li W., et al. & Others Viral Pneumonia Screening on Chest X-rays Using Confidence-Aware Anomaly Detection. IEEE Transactions on Medical Imaging. (2020).

[30] Tuncer T., Ozyurt F., Dogan S. & Subasi A. A novel Covid-19 and pneumonia classification method based on F-transform. Chemometrics And Intelligent Laboratory Systems. 210 pp. 104256 (2021) https://doi.org/10.1016/j.chemolab.2021.104256 PMID: 33531722.

[31] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng ,Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China,JAMA – J. Am. Med. Assoc., 323 (11) (2020), pp. 1061-1069,10.1001/jama.2020.1585.

[32] Chaudhary, P. K. and Pachori, R. B. (2020). Automatic diagnosis of COVID-19 and pneumonia using FBD method. In IEEE (Eds.) International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1-7). IEEE. 10.1109/BIBM49941.2020.9313252.

[33] Ibrahim A., Ozsoz M., Serte S., Al-Turjman F. & Yakoi P. Pneumonia classification using deep learning from chest X-ray images during COVID-19. Cognitive Computation. pp. 1–13 (2021) https://doi.org/10.1007/s12559-020-09787-5 PMID: 33425044.

[34] Zubair S.An Efficient Method to Predict Pneumonia from Chest X-Rays Using Deep Learning Approach. The Importance of Health Informatics in Public Health During a Pandemic. 272 pp. 457 (2020).

[35] Deng J., Dong W., Socher R., Li L., Li K. & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248-255 (2009).

[36] Kundu R., Basak H., Singh P., Ahmadian A., Ferrara M. & Sarkar R. Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. Scientific Reports. 11, 14133 (2021,7), https://doi.org/10.1038/s41598-021-93658-y PMID: 34238992.

[37] Manna A., Kundu R., Kaplun D., Sinitca A. & Sarkar R. A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. Scientific Reports. 11, 1–18 (2021) https://doi.org/10.1038/s41598-021-93783-8 PMID: 34267261.

[38] Jaiswal A., Tiwari P., Kumar S., Gupta D., Khanna A. & Rodrigues J. Identifying pneumonia in chest Xrays: A deep learning approach. Measurement. 145 pp. 511–518 (2019) https://doi.org/10.1016/j. measurement.2019.05.076.

[39] Gabruseva T., Poplavskiy D. & Kalinin A. Deep learning for automatic pneumonia detection. Proceedings Of The IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 350-351 (2020).

[40] Kadry S., Nam Y., Rauf H., Rajinikanth V. & Lawal I. Automated Detection of Brain Abnormality using Deep-Learning-Scheme: A Study. 2021 Seventh International Conference on Bio Signals, Images, And Instrumentation (ICBSII). pp. 1-5 (2021).

Artificial Neural Network Model for Predicting Equivalent Circulating Density of Drilling Fluid

Nuria I. Elhamali
Engineering Project Management Department
School of Applied Sciences and Engineering
Libyan Academy for Postgraduate studies,
Tripoli, Libya
noorlibya947@yahoo.com

Saber Kh. Elmabrouk
Chemical & Petroleum Engineering Department
School of Applied Sciences and Engineering
Libyan Academy for Postgraduate Studies,
Tripoli, Libya
Saber.elmabrouk@academy.edu.ly

***Abstract*— The drilling operation is the most confusing part of the oil and gas industry. The complexity and unexpected problems, as well as the loss of continuity in the drilling process, can result an additional costs and work time. During the drilling, the weight of the drilling fluid must remain above the pore pressure and below the fracture pressure. If the weight of the drilling fluid falls below the pore pressure, there is a risk of gas kicking. In contrast, if it exceeds the fracture pressure, there is a risk of losing the circulation of the drilling fluid. As a result, the equivalent circulating density (ECD) of the drilling fluid is one of the most important key parameters that controls the pressure in the wellbore. Therefore, proper design and management of ECD prior to drilling is key to safe, economical and efficient drilling. The aim of this study is to predict the ECD of the drilling fluid through an artificial neural network (ANNs) model. A total of 872 data points were collected from more than 15 drilled wells to build an ANN model. The selected model consisted of seven inputs, one hidden layer with seventeen neurons, and one output layer. The selected model resulted in the best performance: R2 0.9996 in the training process, 0.9985 in the testing process, 0.5840% average absolute percentage error during training and 0.5651% during testing the model.**

Keywords— equivalent circulating density, artificial neural network, density of drilling fluid, wellbore instability

I. INTRODUCTION

The density of the drilling fluid (mud) regulates the hydrostatic pressure in the drilling, so during drilling the mud weight must remain in a given mud density window above the pore pressure and below the fracture pressure. Drilling fluid is a complex mixture of water-based, oil-based or synthetic-based muds with a number of additives, such as chemical and mineral constituents. If the mud weight drops below the pore pressure, the drill risks kicking where the unexpected and unwanted inflow of reservoir fluid, oil, water, or gas into a borehole due to an unbalanced condition where the pressure within the borehole or bottom hole pressure is less than the formation pressure. In contrast, if the mud weight exceeds the fracture pressure, which is called overbalanced drilling, the hydrostatic pressure of the formation is not high enough to prevent fluid infiltration and circulation losses [1]. Lost circulation means loss of drilling fluid into the formation. Losses can occur if the drilling fluid exceeds the fracture pressure of the formation, causing fractures or penetrating the pore space of the formation.

Equivalent circulation density (ECD) is an important parameter in controlling the pressure in the well during drilling operations. This becomes vital in deep-water drilling where the formation pressure is very close to the fracture pressure (narrow drilling mud density window) as described in Fig. 1. Unaudited ECD may generate a number of problems, such as reduced operational safety, higher pressures, fracturing weak formations, stuck pipe, lost circulation, kicking, blowout, killing operations, and wellbore instability.

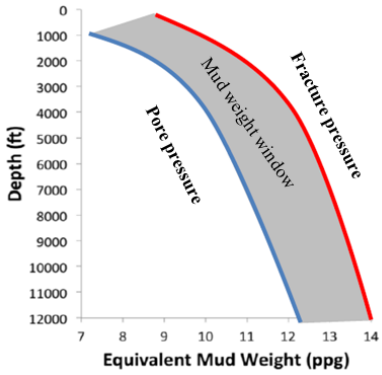


Fig. 1. Mud Weight Window for a Typical Well. Mud Weight Must be Maintained between Pore and Fracture Pressures

During drilling, the mud pump provides pressure to push the mud down into the drill string to the bottom of the hole to exit the drilling bit nozzles and then flow through the annular space between the hole wall and the drill string lifting the rock cuts (drilled cuttings) up to the surface. During this contacting process, some resistance is created and the mud loses the pressure provided by the pump to overcome the frictional resistance caused by the contact. Therefore, the ECD is the sum of this pressure loss (based on density) that complements the original drilling fluid density under static conditions. For this reason, ECD is always more than the density of static mud, as it also includes pressure loss converted to equivalent density. Moreover, drilled cuts can increase mud density and reduce the flow, which in turn increases ECD. Thus, ECD calculations can serve as a basis for real-time monitoring of hole cleaning during drilling. Accordingly, predicting EDC, control, good planning, and best drilling practice are key to preventing drilling problems for safe, economical, and efficient drilling operation. The main goal of this study is therefore to predict ECD using artificial neural network (ANN) techniques that allow us to be ready to properly regulate, control the weight of the mud to stay within a safe range of the hydrostatic pressure, and minimize drilling problems.

In recent years, artificial intelligence (AI) techniques have proven to be an effective tool in solving practical problems in many areas of the O&G industry. The literature shows that ANN, fuzzy logic, and genetic algorithms are the most popular and widely used AI tools in the O&G industry. Some early applications include; fluid flow in pipelines [2], formation damage [3], reservoir characterization [4],

hydraulic fracture design optimization [5], bit selection [6], and oil production prediction [7]. However, the AI, like any other tool, has some limitations. An example of this is ANN, which is often referred to as a black box, only trying to map the relationship between output and input variables according to the training data set. This raises concerns about the ability of the tool to generalize to situations that are not properly included in the data set [8]. With all this, ANN is the most widely used and rapidly evolving. It can solve many problems in optimization, pattern recognition, clustering, time series analysis, forecasting, and control. The main idea behind the creation of ANN is to present a simple model of the human brain to solve complex scientific and industrial problems.

Recently, the O&G industry has been using measurement while drilling and pressure while drilling to predict ECD and monitor changes in bottom hole pressure. These devices include pressure sensors that can measure the bottom hole pressure independently of the factors that control the ECD [9]; [10]; [11]; [12]; [13]. Nevertheless, services companies charge expensive daily rate for these tools, as well as tool operating limitations such as pressure, temperature, tool failures, and so on. These tool costs and rig time can be saved by using AI and machine learning (ML) models as an alternative method to more accurately estimate ECD. Table 1, however, shows some published ECD-AI models. It can be observed that a number of studies have dealt with modeling the ECD, using different AI techniques, with different inputs to predict, estimate and manage ECD before drilling the oil wells. The most interesting and accurate techniques are ANN, LSSVM and ANFIS. It can also be observed that there are different inputs for building the ECD model. However, the robustness of AI models depends on the features selected and the training algorithm used to train the model.

TABLE I. SOME PUBLISHED ECD-AI MODELS

Ref.	Inputs	Data	Model	R^2
[14]	P, T & Initial density	664 points from literature	FIS	0.727
			GA-FIS	0.94
			PSO-ANN	0.996
[15]	P, T & Initial density	-	PSO-ANFIS	0.832
			LSSVM	0.999
			ANFIS	0.850
[16]	MW, DPP & ROP	2376 data points	ANN	0.99
			ANFIS	
[17]	Q, MW, PV, YP, TFA, RPM & WOB	2000 wells	ANN	0.982
[18]	P, T, Mud Type & Initial density	884 points from literature	RBF	0.99
[19]	Q, HL, ROP, RS, SPP, WOB & DT	3567 data points	RF	0.99
			FN	0.99
			SVM	0.97
[20]	GPM, ROP, RPM, SPP, OB & DT	3570 data points	ANN	0.98
			ANFIS	0.96

II. DATA COLLECTION

Drilling data related to ECD was collected from a number of drilling reports, including daily drilling reports, technical reports, mud logging reports, final drilling reports, and drilling bit records from an oilfield in Libya. The outliers were removed from the collected data using box plots. The study then used more than 15 wells drilled in the same area and eventually collected 872 points. Then, the input parameters were selected based on the authors' opinions and experience. Table 2 illustrates the statistics analysis of the seven selected input parameters, which is; weight on bit (WOB), mud weight (MW), revolutions per minute (RPM), plastic viscosity (PV), yield point (YP), flow rate (Flow), and nozzles total flow area (TFA). Fig. 2 shows the correlation matrix between the seven inputs.

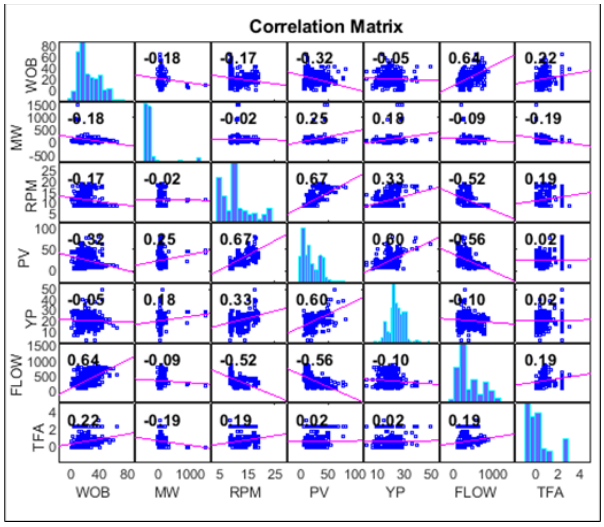


Fig. 2. The Correlation Matrix Between Inputs

In ML, model validation is called a process in which a trained model is evaluated with a set of test data. The test data set is a separate part of the same data set from which the training set is derived. The main purpose of using a test data set is to test the generalizability of a trained model [21]. The model is validated after model training. As shown in Fig. 3, together with model training, the goal of model validation is to find the optimal model with the best performance. Accordingly, the ECD data was divided into 70% for network training and 30% for model testing. Thus, the ANN model was developed using 569 data points and tested with 244 data points. As shown in Fig. 3, the final step in the process of building the ML model is to test the model with an independent dataset. Based on this, 59 datapoints were allocated for the ANN model accuracy study.

III. THE NETWORK ARCHITECTURE

Selecting the optimal network architecture can be achieved with a try error. The most important step, however, is to choose a logarithm that is suitable for creating the network. Accordingly, several algorithms were tested and the algorithm with the highest R2 and the lowest MSE was selected. Table 3 summarizes the algorithms examined in this study. However, the Bayesian Regularization algorithm shows the highest R2 in the training dataset and the test dataset (0.9996 and 0.9985), and the lowest MSE of 0.007 in the training set. However, Error analysis provides an in-depth understanding of the distribution of errors in a model, as well as active data discovery and interpretability techniques for

debugging [31]. Generally, the mean absolute percentage error (MAPE), the coefficient of determination (R2), the mean square error (MSE), and root mean square error (RMSE) are used to test the performance of the models.

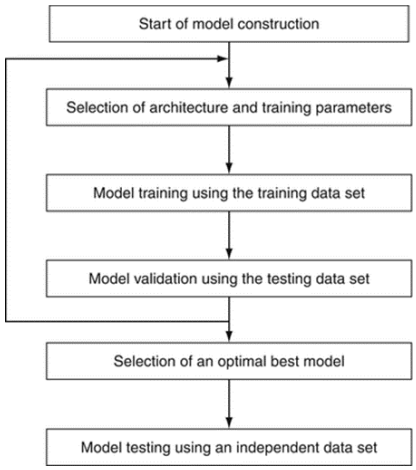


Fig. 3. Illustration of the Process of Building a ML Model, after [21]

TABLE II. STATISTICS ANALYSIS OF THE INPUT PARAMETERS

	WOB	RPM	MW	PV	YP	Flow	TFA
Max	65	1500	18.9	82	50	900	3.142
Min	1	10	8.6	5	2	95	0
Mean	22.06888	144.663	11.68181	25.30913	20.52891	355.8604	0.699619
Standard	10.99844	243.9034	2.843682	14.53068	5.775359	201.9629	0.713216
Variation	120.9657	5948 8.86	8.086527	211.1407	33.35477	4078 9.02	0.508677

However, MAPE is a measure of how accurate a forecast model is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. R2 measures the rate of variation in the dependent variable that can be attributed to the independent variable. The value of R2 is always between 0 and 1. While, MSE measures the degree of error in the models. Assess the mean squared difference between observed and predicted values. If there is no error in the model, the MSE is zero. As the model error increases, its value increases. Another way to assess how well a model fits into a dataset is to RMSE, which is a measure the average distance between the predicted values from the model and the actual values in the dataset.

TABLE III. ALGORITHMS EXAMINED IN THIS STUDY

Algorithm	MSE		R^2	
	Train	*Test*	*Train*	*Test*
Bayesian Regularization	0.007	0.006	0.9996	0.9985
Levenberg-Marquardt	0.007	0.016	0.9989	0.9987
BFGS Quasi-Newton	0.017	0.025	0.9983	0.9988
Resilient Backpropagation	0.12	0.11	0.991	0.9907
Scaled Conjugate Gradient	0.05	0.049	0.9958	0.9963
Conjugate Gradient with Powell/Beale Restarts	0.035	0.34	0.9972	0.9973
Fletcher-Powell Conjugate Gradient	0.041	0.042	0.9966	0.9967
Polak-Ribiére Conjugate Gradient	0.052	0.051	0.9957	0.9965
One Step Secant	0.054	0.055	0.9956	0.9958
Variable Learning Rate Gradient Descent	0.13	0.14	0.9895	0.9903

The number of layers and nodes in each hidden layer was optimized based on the iterative process. The results of this process are shown in Table 4 and 5. Obviosity, one hidden layer gives highest R2 in training and testing dataset (0.9996 and 0.9985) and lowest MSE of 0.006 in testing dataset. On the other hand, seventeen nodes in the hidden layer gives highest R2 and lowest MSE.

TABLE IV. DIFFERENT NUMBER OF HIDDEN LAYERS

Hidden Layer	**Architecture**	**MSE**		**R^2**	
		Train	*Test*	*Train*	*Test*
1	[7-17-1]	0.007	0.006	0.9996	0.9985
2	[7-17-20-1]	0.010	0.124	0.9992	0.9969
3	[7-17-20-10-1]	0.022	0.285	0.9970	0.9949
4	[7-17-20-10-15-1]	6.450	5.887	0	0
5	[7-17-20-10-15-10-1]	6.946	5.762	0	0

TABLE V. DIFFERENT NUMBER OF NEURONS IN THE HIDDEN LAYER

Neurons	**MSE**		**R^2**	
	Train	*Test*	*Train*	*Test*
10	0.011	0.009	0.9994	0.9970
15	0.012	0.008	0.9995	0.9970
17	0.007	0.006	0.9996	0.9985
20	0.011	0.017	0.9996	0.9987
30	0.011	0.026	0.9998	0.9959

Fig. 4 shows the training performance. It shows the MSE of the Bayesian Regularization algorithm for training and testing. This training stops after 384 iterations at which the training set MSE is minimal.

With this step, a supervised neural network was created with one input layer with seven nodes, one hidden layer with seventeen neurons, and one output layer, [7-17-1], to predict ECD, as shown in fig.5. However, the training function is the overall algorithm that is used to train the neural network to recognize a certain input and map it to an output. The ANN model trained by Trainbr function with Bayesian Regularization algorithm. Whereas, Trainbr is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then

determines the correct combination so as to produce a network that generalizes well. The process is called Bayesian regularization.

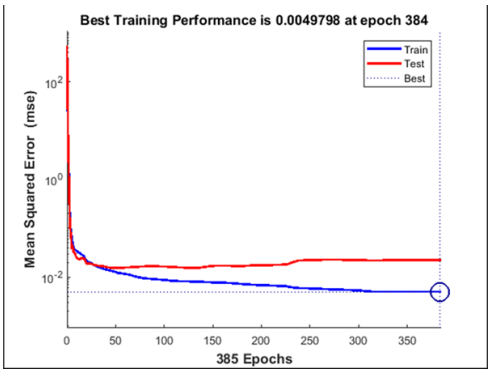


Fig. 4. MSE vs Epochs for the Bayesian Regularization Training Function

For the hidden layer, tan-sigmoid transfer function was selected, and the linear transfer function was chosen for the output layer. Table 6 illustrates the characterization of the selected ANN model to predict ECD.

TABLE VI. THE BEST MODEL SELECTION

Architecture		[7-17-1]
Neurons		17
Training Function		Trainbr
Algorithm		Bayesian Regularization
Transfer function for the hidden layer		Tan-sigmoid
Transfer function for the output layer		Linear
MSE	Train	0.007
	Test	0.006
R^2	Train	0.9996
	Test	0.9985
MAPE	Train	0.58
	Test	0.56
Min Error	Train	0.00
	Test	0.01
Max Error	Train	5.92
	Test	4.22

IV. RESULTS AND DISCUSSION

Fig. 5 shows the actual and predicted ECD used for training, testing, and overall. In the diagrams, most of the data points are scattered too close to the Y = T line, indicating that the selected ANN model provides very good performance. As a result of the network, R2 was obtained at 0.99964 for training, 0.9985 for testing, and 0.99946 for overall data. As other indications of ECD ANN model performance, Fig. 6 shows the error histogram of the Bayesian regularization training algorithm. The histogram shows the normal distribution of errors. Blue bars indicate training data and red bars testing data. Most errors in training and testing data points are close to zero and no outliers are seen. Therefore, the model provides very good performance.

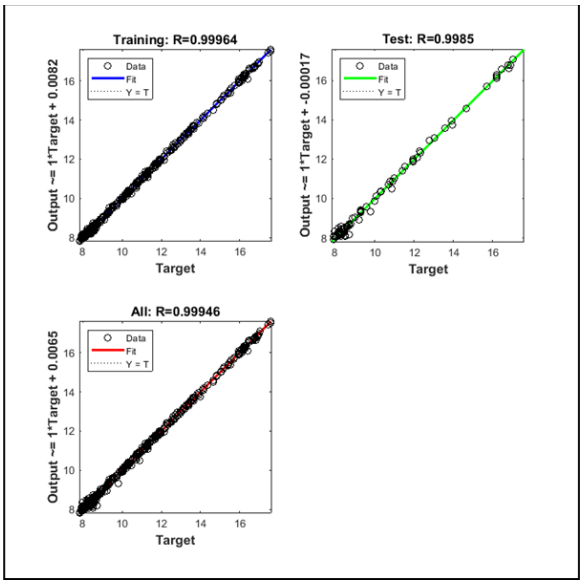


Fig. 5. Predicted and Actual ECD for Training, Testing, and All Datasets

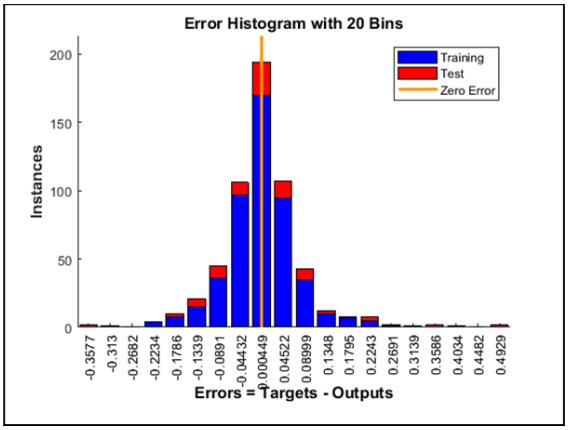


Fig. 6. Error Histogram for the Training Algorithm

Consequently, the ECD empirical correlation based on ANN is given by the Equation 1.

$$\mathrm{ECDn} = \left[\sum_{i=1}^{N} w_{2i}\left(\cdot\frac{2}{1+e^{-2(\sum_{f=1}^{F}(w_{1i,f}x_f+b_{1i})}}-1\right)\right]+b_2 \quad (1)$$

Where, ECDn is the normalized ECD, b1 is the bias of the hidden layer, b2 is the bias of the output layer, N is the number of neurons in the hidden layer that was optimized to be 17, w1 is the weight of the hidden layer, w2 is the weight of the output layer, and x is the inputs. Also, f1 is WOB, f 2 is RPM, f3 is MW, f 4 is PV, f 5 is YP, f 6 is Flow, and f 7 is TFA. All biases and weights related to the equation are shown in Table 7. Then, ECD can be calculated from ECDn using the Equation 2.

$$\mathrm{ECD} = ((\mathrm{ECDn}+1)*4.87)+7.82 \quad (2)$$

TABLE VII. WEIGHTS AND BIASES OF THE ECD CORRELATION (EQ. 1)

HL neurons, N	w between IPL & HL, w_{1f}							HL biases	w between OPL & HL	OPL bias
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	b_1	w_2	b_2
1	0.279097	-0.07765	-1.53143	0.118202	0.982502	-0.01983	0.563947	-0.00065	-0.78243	
2	0.600134	-0.15494	-2.66331	-0.73737	0.598639	0.629102	0.0406	-0.96939	-0.83558	
3	0.095074	-0.2207	0.420996	-0.02532	1.127482	-0.30823	-0.01639	0.826089	1.597031	
4	-0.94478	0.233632	-0.01364	0.813472	-0.54009	-0.31656	-0.23674	-0.01773	1.011377	
5	0.599253	-0.26051	-0.02513	0.337307	0.439505	-0.43697	0.490908	-0.16739	-0.8179	
6	0.509267	0.238948	1.417334	-0.5634	0.565545	2.037209	-0.09171	1.084027	0.22128	
7	0.792625	-0.12004	-0.09978	-0.20546	-0.67346	1.305134	-0.57349	-0.39815	0.959235	
8	-0.5718	-0.05239	0.178043	0.444707	0.304494	-1.89552	0.503602	0.558135	0.764693	
9	-0.98443	0.20761	-0.03411	0.90834	-0.47331	-0.4183	-0.84733	-0.1669	-0.4552	-0.01269
10	-0.49593	0.402477	-1.38533	-0.56365	0.352016	0.141686	0.563945	-0.01842	1.044079	
11	-0.56829	0.386221	0.443836	-0.83363	-0.756	-0.35289	0.194259	-1.86763	0.712096	
12	0.685878	-0.16628	-1.54072	-0.13986	0.489155	0.276271	0.086626	-0.29767	2.197246	
13	0.066311	-0.21827	1.710647	0.554376	-0.33044	-0.36497	-0.37727	-0.31868	0.911821	
14	-0.13053	-0.46535	0.511491	-0.89118	1.832328	-0.48081	0.154172	0.288747	-0.73798	
15	-0.83241	0.316246	-0.46076	-0.32445	-0.45189	0.461527	-0.24351	0.725478	-1.06857	
16	0.623654	0.006096	0.682744	0.083071	-0.58269	0.095791	0.127976	0.671971	0.856098	
17	0.93337	0.359564	-0.52061	0.77167	-0.77865	0.525812	-0.4668	0.363205	-0.82878	

V. MODEL ACCURACY TEST

As explained above, the final step in building the ML model is to test the accuracy of the proposed ANN model with an independent data set. Therefore, use the independent data set (59 datapoints) as the input to Equations 1 and 2 to calculate the ECD using the weights and distortions shown in Table 7. Note that before applying Equations 1 and 2, the input data must be normalize between -1 and 1 by Equation 3.

$$x' = 2 * \left(\frac{x - xmin}{xmax - xmin}\right) - 1 \qquad (3)$$

Where, x' is the normalized value, x is the input, xmax and xmin are the maximum and the minimum of x respectively.

Fig. 7 shows the proposed ANN model accuracy. It is clear that the points are too close scattered around the y = x line, indicating that the model predicts the ECD with a high accuracy of 0.9966 R2. However, with this way, the calculated ECD value with the resulting empirical correlation (Eq 1) allows us to be ready to properly regulate and control the mud to stay within the safe range of hydrostatic pressure, manage and minimize drilling problems prior to drilling the oil wells.

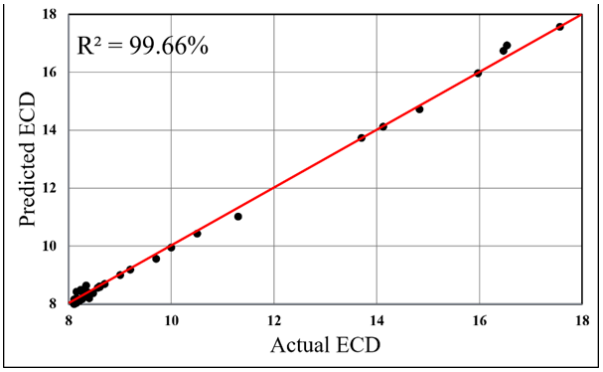


Fig. 7. Calculated ECD vs. Actual Values

VI. CONCLUSION

ECD was predicted using an ANN whit 813 data points from 15 wells collected from the same field in Libya. The data were used to construct a model with a 70:30 train-to-test ratio dataset. The ANN input includes; weight on bit, revolutions per minute, mud weight, plastic viscosity, yield point, flow rate, and nozzles total flow area.

The proposed ANN model was constructed using one input layer with 7 nodes, one hidden layer with 17 neurons, and one output layer. The transmission function of the hidden layer was tan-sigmoid, and the algorithm was BR. The R2 is 0.9996 in training, 0.9985 in testing, and 0.9994 in all datapoints. The MAPE is 0.5840% in training and 0.5651% in testing. This indicates that the AND technique has a greater ability to predict ECD.

The resulting empirical correlation of ECD (Equation 1) was applied to another data (59 points) and a good predicted results were obtained with 0.9966 R2, indicating that the proposed ANN model is a robust, and high-precision model.

ACKNOWLEDGMENT

REFERENCES

[1] Savari, S., Whitfil, D., Jamison, D.E., Kumar, A. 2014. A method to evaluate lost circulation materials—Investigation of effective wellbore strengthening applications. SPE 167977.

[2] Alhashem, M. (2020, January). Machine Learning Classification Model for Multiphase Flow Regimes in Horizontal Pipes. In International Petroleum Technology Conference. OnePetro.

[3] Effiong, Augustine James, Etim, Joseph Okon, and Anietie Ndarake Okon. "Artificial Intelligence Model for Predicting Formation Damage in Oil and Gas Wells." Paper presented at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 2021. doi: https://doi.org/10.2118/207129-MS

[4] Okon, Edet Ita, and Dulu Appah. "Application of Machine Learning Techniques in Reservoir Characterization." Paper presented at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 2021.

[5] Morozov, A., Popkov, D., Duplyakov, V., Mutalova, R., Osiptsov, A., Vainshtein, A., ... & Paderin, G. (2020, November). Machine Learning on Field Data for Hydraulic Fracturing Design Optimization: Digital Database and Production Forecast Model. In First EAGE Digitalization Conference and Exhibition (Vol. 2020, No. 1, pp. 1-5). European Association of Geoscientists & Engineers.

[6] Batruny, Peter , Zubir, Hafiz , Slagel, Pete , Yahya, Hanif , Zakaria, Zahid , and Alaa Ahmad. "Drilling in the Digital Age: Machine Learning Assisted Bit Selection and Optimization." Paper presented at the International Petroleum Technology Conference, Virtual, March 2021.

[7] Elmabrouk, S., Shirif, E., & Mayorga, R. J. P. S. (2014). Artificial neural network modeling for the prediction of oil production. Petroleum Science and Technology, 32(9), 1123-1130.

[8] J. Brownlee, "Neural Networks are Function Approximation Algorithms," Machine Learning Mastery, Mar. 17, 2020. https://machinelearningmastery.com/neural-networks-are-function-approximators/.

[9] Rommetveit, Rolv, and Christine Nordstrand. "Experiences from use of a Managed Pressure Drilling & ECD Management System during drilling of a challenging HPHT well." SPE/IADC Managed Pressure Drilling and Underbalanced Operations Conference and Exhibition. Society of Petroleum Engineers, 2010.

[10] Andagoya Carrillo, K. I., Avellán, F. J., & Camacho, G. (2015, November). ECD and downhole pressure monitoring while drilling at ecuador operations. In SPE Latin American and Caribbean Petroleum Engineering Conference. OnePetro.

[11] Elzenary, M., Elkatatny, S., Abdelgawad, K. Z., Abdulraheem, A., Mahmoud, M., & Al-Shehri, D. (2018, April). New technology to evaluate equivalent circulating density while drilling using artificial intelligence. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. OnePetro.

[12] Alsaihati, A., Elkatatny, S., & Abdulraheem, A. (2020). Real-time prediction of equivalent circulation density for horizontal wells using intelligent machines. ACS omega, 6(1), 934-942.

[13] AlSaihati, A., Elkatatny, S., Gamal, H., & Abdulraheem, A. (2021, May). A Statistical Machine Learning Model to Predict Equivalent Circulation Density ECD while Drilling, Based on Principal Components Analysis PCA. In SPE/IADC Middle East Drilling Technology Conference and Exhibition. OnePetro.

[14] Ahmadi, M. A., Shadizadeh, S. R., Shah, K., & Bahadori, A. (2018). An accurate model to predict drilling fluid density at wellbore conditions. Egyptian Journal of Petroleum, 27(1), 1-10.

[15] Ahmadi M. A. Toward Reliable Model for Prediction Drilling Fluid Density at Wellbore Conditions: A LSSVM Model. Neurocomputing 2016, 211, 143–149. 10.1016/j.neucom.2016.01.106.

[16] Abdelgawad K. Z.; Elzenary M.; Elkatatny S.; Mahmoud M.; Abdulraheem A.; Patil S. New Approach to Evaluate the Equivalent Circulating Density (ECD) Using Artificial Intelligence Techniques. J. Pet. Explor. Prod. Technol. 2019, 9, 1569–1578. 10.1007/s13202-018-0572-y.

[17] Alkinani, Husam H., et al. "Data-Driven Neural Network Model to Predict Equivalent Circulation Density ECD." SPE Gas & Oil Technology Showcase and Conference. OnePetro, 2019.

[18] Rahmati A. S.; Tatar A. Application of Radial Basis Function (RBF) Neural Networks to Estimate Oil Field Drilling Fluid Density at Elevated Pressures and Temperatures. Oil Gas Sci. Technol. 2019, 74, 50.10.2516/ogst/2019021.

[19] Alsaihati A.; Elkatatny S.; Abdulraheem A. Real-Time Prediction of Equivalent Circulation Density for Horizontal Wells Using Intelligent Machines. ACS Omega 2021, 6, 934–942. 10.1021/ACSOMEGA.0C05570

[20] Gamal, H., Abdelaal, A., & Elkatatny, S. (2021). Machine Learning Models for Equivalent Circulating Density Prediction from Drilling Data. ACS omega, 6(41), 27430–27442. https://doi.org/10.1021/acsomega.1c04363

[21] Wang, H., & Zheng, H. (2013). Model Validation, Machine Learning. Encyclopedia of Systems Biology, 1406–1407. doi:10.1007/978-1-4419-9863-7_23

Hybrid Compression-Watermarking Scheme of Fetal Heart audio Signal for Telemedicine Application

Ali Abdrhman Ukasha
Electrical and Electronic Engineering Depatment
Sebha University
Sebha, Libya
ali.ukasha@sebhau.edu.ly

Nassir Mansour Abuhamoud
Electrical and Electronic Engineering Depatment
Sebha University
Sebha, Libya
mans.abuhamoud1@sebhau.edu.ly

Mohamed Omar Banana
Electrical and Electronic Engineering Depatment
Sebha University
Sebha, Libya
moh.babana@sebhau.edu.ly

***Abstract*— One of the important bio-electrical signal is Electrocardiogram (ECG) measured after a certain time, in addition to that is a very important diagnosis device in clinical application. This work introduces useful approach for Fetal ECG signal audio file compression and ECG watermarking for telemedicine applications. For embedding binary watermark image data in ECG signal, the Least Significant Bit (LSB) watermarking algorithm is used. The objective of this paper is to develop an efficient algorithm for ECG compression & embedding contour points instead of binary watermark image. The contour image is extracted using single step parallel contour extraction (SSPCE) method. The first stage is Fetal ECG audio signal compression process: First, the ECG signal is compressed using discrete cosine transform (DCT) by selecting the resulting transformed coefficients by through thresholding process. Then these coefficients are coded using the run-length encoding (RLE) scheme. The second stage is embedding binary watermark image data in the coded ECG signal by using LSB watermarking algorithm. In the receiver side the contour points are extracted from LSB of the audio ECG signal file. Simulation results show that the Compression ratio (CR) is exceeds to 95%, percentage root mean error difference (PRD) is 0.1425, quality score (QS) is 677.2070, and the normalized correlation coefficient (NCC) closest to 1.00. By these results, good improvement in the compression ratio and quality of extracted watermark is obtained in case of using hybrid compression-watermarking.**

Keywords—Fetal ECG Signals; ECG compression; DCT transform; RLE encoding; Signal reconstruction

I. INTRODUCTION

Embedding information in a host signal is called watermarking. One of the important watermark embedding algorithms out is the Least Significant Bit (LSB). The choice of host signal and volume of embedding information determines the performance of each of the watermark embedding algorithms. The host signal is distorted when embedding data is included. Measuring the distortion between the host and the watermarked signal is one of the most important metrics for evaluating the performance of watermarking algorithms. Less distortion, better performance.Several applications play an important role in ECG data compression such as telemedicine, mobile ECG monitoring, wearable healthcare devices and hospital patient databases [1-3]. The main purpose of the compression technique is to reduce the data area without distorting the morphological features of the signal during reconstruction; the bits for each sample of the ECG signal are lowered in such a way as to ensure that they are correctly represented on the receiver side. Generally, the compression techniques is classified in to two types: lossless compression (which is used in this work) that is leads to low compression ratio in which the original data is reconstructed perfectly without any distortion in the original data; and lossy compression that is has high compression ratio that leads to non-perfect reconstruction from some parts of the original data [4]. Recently some researcher certified that the heart disease is one of the most common causes of death worldwide [5].Therefore, telemedicine services and applications have become so popular in the last decade [6,7]. This work has been implemented using MATLAB. Therefore, ECG signal data compression is a very complex problem and we need a compromise between transmitting its data with a smaller bandwidth while maintaining high efficiency for the purpose of storing it in less space. Instead of inserting a binary watermark, the contour points of the binary image that are inserted into the cover image will be extracted. This leads us to include fewer bits if the extraction process is done with high quality on the future side, and is considered one of the most important challenges for this work in this paper. When inserting a watermark with a large size of pixels, the biggest challenge we face is how to extract that watermark without distortion. This work presents a proposal to send the image of the watermark border points instead of sending the image of the digital watermark. At the future, the morphology process is applied to the image of the extracted image boundary to obtain the

recombined binary image. By this way we achieved an abundance of the number of inserted bits while maintaining the high quality of the extracted binary image.

II. CONTOUR EXTRACTION USING ''SSPCE'' METHOD

To detect if an edge is present or not, we pass a 3×3 filter across an image (i.e. pixel by pixel). An eight-way Freeman string notation is used [20]. SSPCE method is explained as follows [9], [11]: When the central pixel of a window is connected to any of its eight surrounding neighbors, then we have a detected edge. A window within an object's region will be considered to be object pixels if all of the object's pixels (here: nine pixels) are facing nine pixels by an edge extraction rule, and vice versa occurs when all nine pixels of the object facing the nine pixels of the edge extraction base are background pixels, so the window will be within the object's region. If the central pixel of an object is surrounded by pixels in the background, the central pixel is considered as a noise point and this is due to the digitization of the image [9], [10] and [11]. Compressing the ECG signal is a challenging task due to the randomness of the ECG signal. Therefore, it is difficult to achieve a high rate using uncompromised compression techniques [16]. The block diagram of the contour extraction points is shown in Fig. 1.

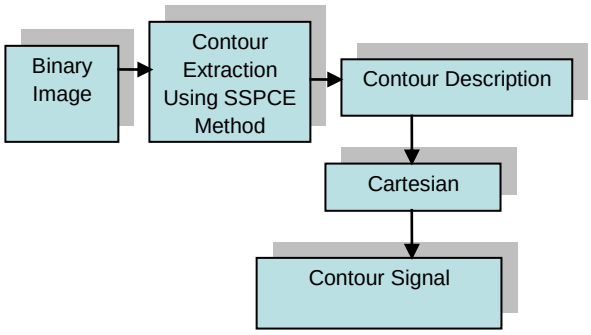


Fig. 1. Contour extraction representation

III. DISCRETE COSINE TRANSFORM (DCT)

In image processing applications DCT is a popular technique for compression purposes. The lossy image compression offered by DCT results in a better quality image with a high compression ratio. Due to its high decorrelation and energy compaction property; DCT is a basis for many signal and image compression algorithms. Transformations generally mean moving usually from the original (spatial) field to another (frequency) and vice versa. The DCT Transform is the most common and widely used for many digital signal processing applications. Two of its main properties of this transform: energy conservation; and concentration in several parameters close to zero frequency made this transform to give us good performance, where it was used in the JPEG standard for image compression. 2D forward and inverse discrete Cosine Transform of *n* sample is defined as in equations 1 & 2 respectively [17]:

$$F(u,v)=\frac{4C(u)C(v)}{n^2}\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}f(j,k)\cos[\frac{(2j+1)u\pi}{2n}]\cos[\frac{(2k+1)v\pi}{2n}] \quad (1)$$

$$f(j,k)=\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}C(u)C(v)F(u,v)\cos[\frac{(2j+1)u\pi}{2n}]\cos[\frac{(2k+1)v\pi}{2n}] \quad (2)$$

where:

f(j,k) is the host image (size *n* x *n*) in spatial domain *(j,k); F(u,v)* is the image in spectrum domain *(u,v); C (u), C (v)* = $1/\sqrt{2}$ for *u, v*=0 and *C (u)* , *C (v)* =1 otherwise.

IV. BIT-PLANE DECOMPOSITION (LSB)

We assume a 256 X 256 medical pixels image to be given in 8bit/pixel (bpp) precision. The entire image is considered as a two-dimensional array of pixel values. We consider the 8bpp data in the form of 8-bit planes, each bit plane associated with a position in the binary representation of the pixels. 8-bit data is a set of 8 bit-planes. Each bit-plane may have a value of 0 or 1 at each pixel, but together all the bit-planes makeup a byte with value between 0 to 255.

One of the most important processing tools in digital image processing is the representation of the bit-plane. The bit-plane is a set of digital numbers used to represent the bit value in each plane. For 8-bit grayscale images, there are eight different binary images that is obtained to analyze the image into eight image planes. For each bit in each plane there is a probability of 0, 1 values. The Least Significant Bit (LSB) is the zero bit plane of the last bit of each screen element value of the representation in binary form. On the other hand, the most significant bit (MSB) refers to bit plan $(n-1)^{th}$ for each value of a screen element in the image in binary representation. In the case of 8-bit color images, there are 8 binary images that is obtained for each color channel. Fig. 2 shows a gray image with corresponding bit planes. Equation (3) is used to obtain a gray level from bit planes values [13].

$$gray\ level(x,y)=\beta_0 \times 2^0+\beta_1 \times 2^1+\beta_2 \times 2^2+\cdots+\beta_7 \times 2^7 \quad (3)$$

where: β_i is bit-plane information at i^{th} plane, *i*=[0,1,2,3, …,7].

In the higher bit planes, the distribution has irregular due to alignment between zeros and ones. That is why the low bit planes of grey images is very common because it has a uniform distribution. This type of bias is known as a redundancy and it progresses gradually as we move from the lower bit plane to higher bit plane. So, bits in one or more of the higher bit planes is compressed to leave space to hide data as a watermark. The frequency band is more appropriate than the time domain to obtain a large bias between zeros and ones [12].

V. EMBEDDING/EXTRACTION PROCEDURES

The following steps are sequentially used for the embedding/extraction of the secret watermark image within the cover image in the proposed watermarking algorithm:

1. Patient's Fetal ECG audio data is selected as host signal for watermarking.
2. ECG audio file is compressed using DCT & RLE.
3. Contour points is extracted using SSPCE method from original binary watermark.
4. Apply LSB data embedding algorithm; to Embedding contour data into ECG audio compressed file in the least significant bit.
5. Random Permutation (data encryption method by shuffling pixel information randomly within the image) for data Pixels up to 80% from total pixels in the image.
6. In the receiver side; the contours points are extracted.
7. Apply some morphological operations to obtain the reconstructed binary watermark image.
8. Evaluate performance of the proposed system by MSE, PSNR, PRD, CR, NCC, and QS.

The block diagram of the embedding/extraction steps are shown in Fig. 2.

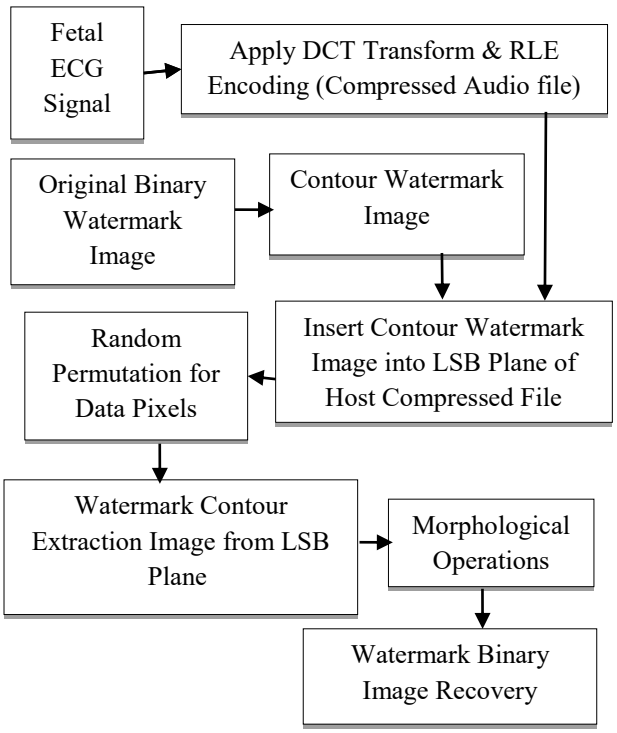


Fig. 2. Embedding/Extraction Watermark Binary Image

VI. ENCODING ALGORITHM

The purpose of ECG signal compression is for confidentiality of important information and to obtain a suitable size for it, as well as to obtain a high CR compression ratio while maintaining the accuracy of the signal. One of the telemedicine is ECG signals, therefore we need to compress it and transmit it.

The run-length is a lossless encoding method that takes advantage of the consecutive repetitions of a specific number. Because the thresholding step sets to zero a large number of coefficients, the run-length scheme represents the data by the zero follows by the total of repetitions. If the coefficient is different to zero, the encoded data is equal to the wavelet coefficient. In our architecture, the output of the system is data and row; data is the encoded wavelet coefficient and row is the position into the run-length code.

This is a very simple compression technique method used for compressing sequential data. Many digital image consist pixel values that are repeats sequentially for such type of image RLE is useful. For more information can be found in [18, 19].

VII. RESULTS AND STATISTICAL ANALYSIS

The compression ratio (CR) is defined as the ratio of the number of bits representing the original signal to the number of bits required to store the compressed signal. Data compression ratio is a measurement of the relative reduction in size of data representation produced by a data compression algorithm. It is typically expressed as the division of uncompressed size by compressed size.

$$RC = \frac{number\ of\ bits\ for\ original\ signal}{number\ of\ bits\ for\ Compressed\ signal} \quad (4)$$

A high compression ratio is typically desired. A data compression algorithm must also represent the data with acceptable fidelity while achieving high CR.

The compression ratio of the analyzed method is measured using the equation (5).

$$CR = \frac{Original\ file\ size}{Compressed\ file\ size} \quad (5)$$

The globally standard of ECG signal reconstruction quality in terms of percentage root mean difference PRD is defined as PRD < 2%: very good reconstruction, PRD <2-5%: good reconstruction, PRD < 5-9%: acceptable quality of reconstruction and PRD > 9%: unacceptable quality. PRD calculation is as follows:

$$PRD = 100 * \sqrt{\frac{\sum_{i=1}^{n}(ORG(i)-REC(i))^2}{\sum_{i=1}^{n}(ORG(i))^2}} \quad (6)$$

Where ORG is the original signal, REC is the reconstructed signal and *n* is the length of the window over which the PRD is

calculated. The lower the PRD, the closer the reconstructed signal is to the original ECG data.

Because the PRD in the entire works is not ever in the same range, a parameter that helps to compare the tradeoff between the CR and the PRD is the Quality Score (QS) [15]. This is the relation between the CR and the PRD, represented as:

$$QS = \frac{CR}{PRD} \tag{7}$$

The higher QS, the hiquer relationship between the CR and the PRD.

The normalized cross-correlation (NCC), usually its 2D version, is routinely encountered in template matching algorithms, such as in facial recognition, motion-tracking, registration in medical imaging, etc.

There are various ways to evaluate how much correlation between the extracted watermark and the original watermark. The normalized cross correlation (*NCC*) is one of the well-known measurements that used for this purpose. The normalized cross correlation defines as in equation (8) [14].

$$NCC(w, w^*) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j)w^*(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j)^2} \tag{8}$$

Where *w(i, j)* and *w*(i, j)* are the original and extracted watermarks, respectively.

Quality measuring of an approximation during the approximating procedure uses mean square error (*MSE*) and signal-to-noise ratio (*SNR*) criterions by the equations (9) and (10) respectively.

$$MSE = \frac{1}{L_{CC}} \sum_{i=1}^{L_{CC}} d_i^2 \tag{9}$$

where d_i is the perpendicular distance between *i* point on the curve segment and straight line between each two successive vertices of that segment.

$$SNR = -10 * \log_{10}\left(\frac{MSE}{VAR}\right) \tag{10}$$

where *VAR* is the variance of the input sequence.

To evaluate the proposed method, three binary image of the watermark image are tested which are shown in Fig. 3; where PN is the points number.

Some parameters are obtained during experiments are done as shown in Table I. The compression stage steps are shown in Fig. 4 & Fig.5. The results of the experiments are shown from Fig 6 to Fig 9 (see Tables I, II, and III).

Fig. 3. Tested Binary Images with Size (95x95): a) ''Iqra'' Word, b) ''Libya''Name, and c) ''Sebha University'' Name

TABLE I. MEASURES RESULTS (TH = 0.002 & RP = 80000)

PRD	RC	Compression Ratio (CR %)	QS
0.1425	28.6478	96.5107	677.2070
	PSNR [dB] (Org & Compressed)	PSNR [dB] (Org & Noisy Watermarked)	
	45.9445	16.5813	

TH: Threshold value for DCT coefficients & RP: Random Permutations for data pixels

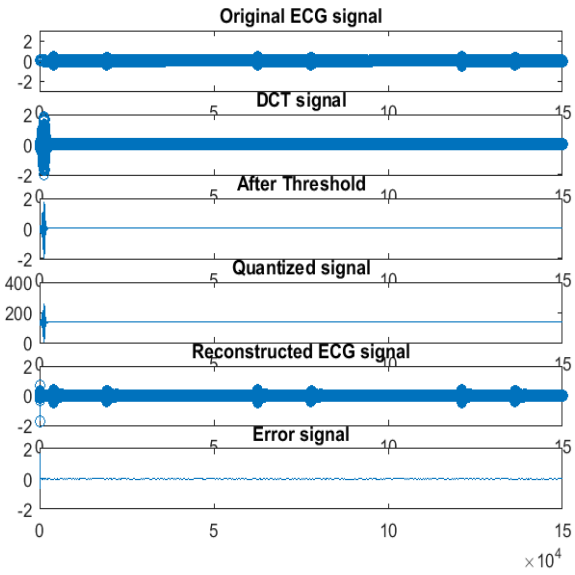


Fig. 4. Compression Stage using DCT transform, Quantization, & RLE Encoding

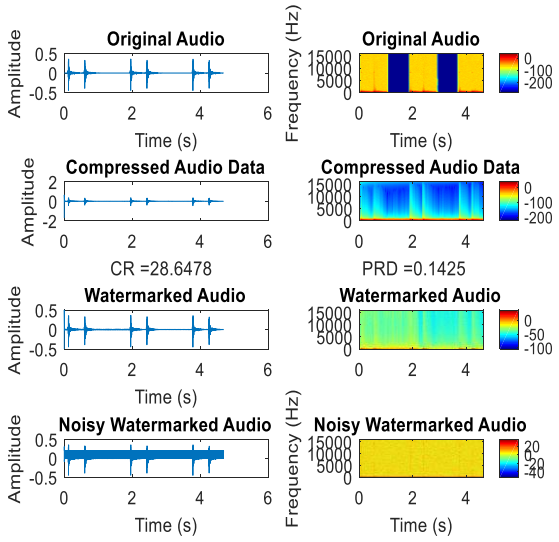


Fig. 5. Original Signal, Compressed, Watermarked, and Noisy Watermarked Audio Signals, and It's Spectrograms

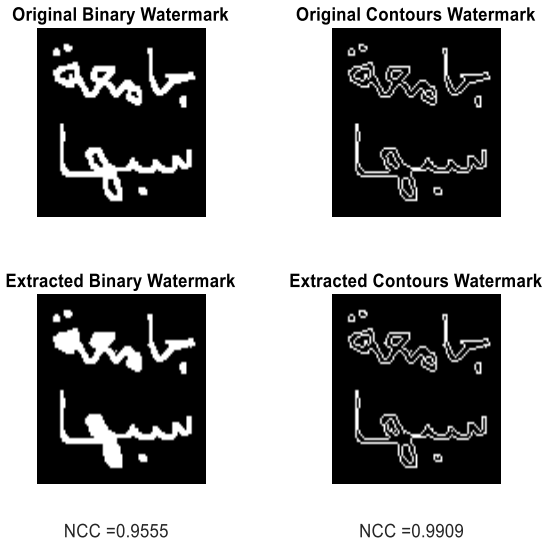


Fig. 6. Results of Extracted Tested Watermark Image with noisy pixels = 55.33% from Compressed-Watermarked Signal

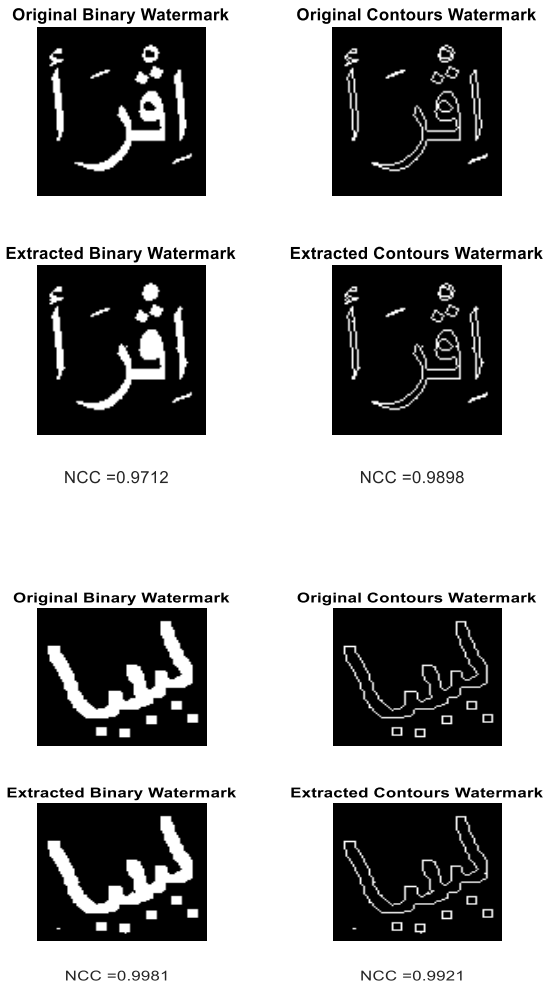
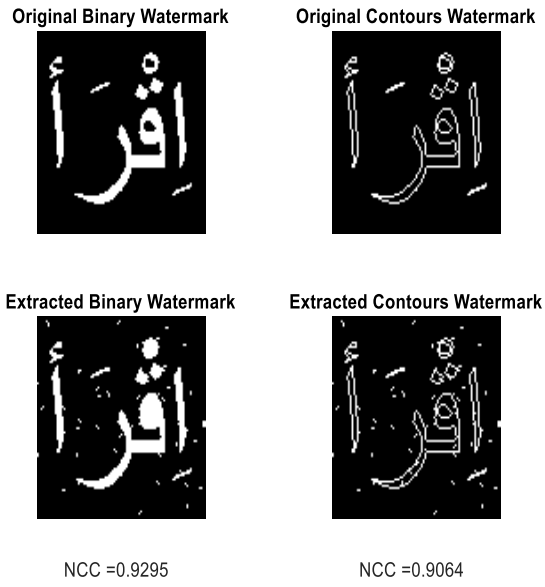


TABLE II. MEASURES RESULTS (TH = 0.002 & DP = 100000)

PRD	RC	Compression Ratio (CR %)	QS
0.1425	28.6478	96.5107	677.2070
	PSNR [dB] (Org & Compressed)	PSNR [dB] (Org & Noisy Watermarked)	
	45.9445	15.6107	



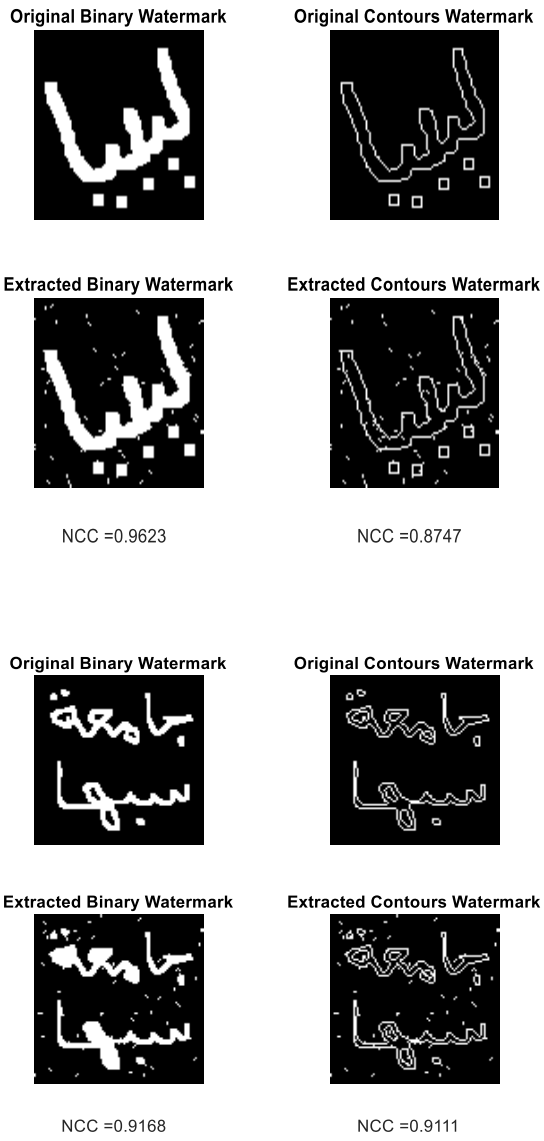


Fig. 7. Results of Extracted Tested Watermark Image with noisy pixels = 66.67% from Compressed-Watermarked Signal

TABLE III. MEASURES RESULTS (TH = 0.002 & OSL = 150000)

PSNR [dB]	No. of Noisy Pixels
16.5813	80000
15.6107	100000
14.8214	120000

OSL: Original signal length

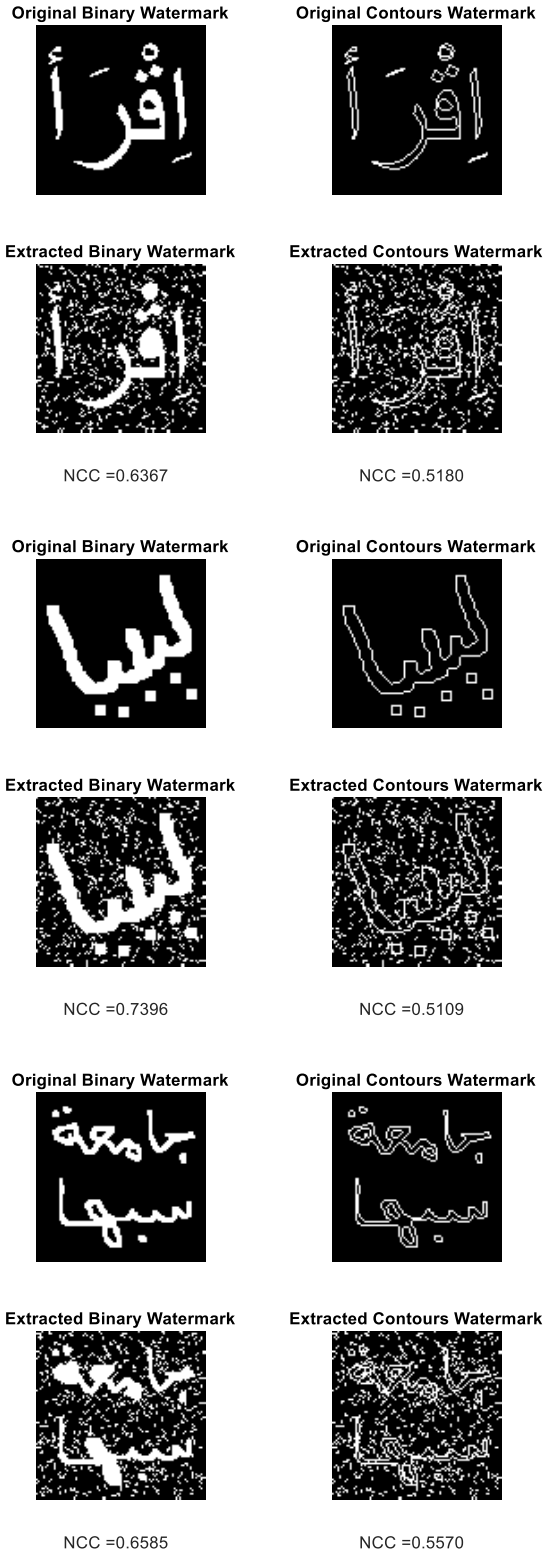


Fig. 8. Results of Extracted Tested Watermark Image with noisy pixels = 80% from Compressed-Watermarked Signal

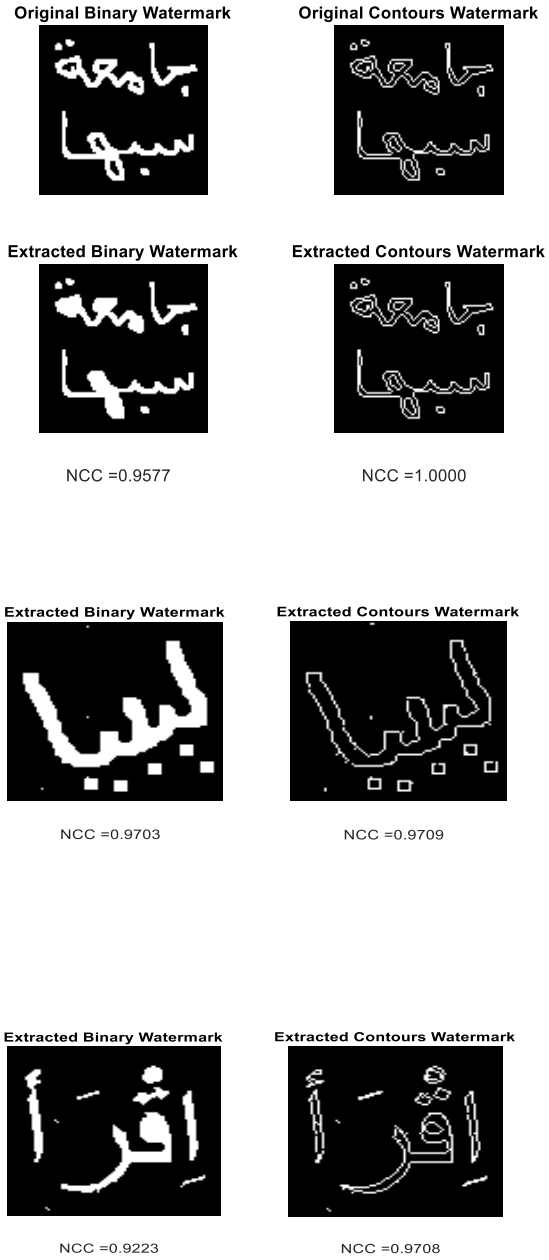


Fig. 9. Results of Extracted Tested Watermark Image with AWGN (33 dB) from Compressed-Watermarked Signal

VIII. Discussion & Conclusions

In this work, a two-stage hybrid ECG signal scheme was proposed: the first stage is to compress the Fetal ECG audio signal using a DCT transform and an RLE encoder, while maintaining the quality of the compressed signal. Some paprameters are obtained in the validity range and with good quality as: compression ratio obtained is exceeds to 95%, percentage root mean error difference (PRD) is 0.1425, and the quality score (QS) is 677.2070. The second stage is the process of inserting the watermark using the LSB level. To reduce the number of insert bits, we use the contour points of the watermark instead of the binary watermark in the insert process. The contour is extracted uses SSPCE contour extraction method. At the receiver side the contour points of the watermark are extracted from LSB plane of the received file audio signal. Finally, some morphological processes are applied to obtain the binary watermark image, where has normalized correlation coefficient varied between 0.998 to 0.636 for the random permutations of data pixels (55.33% up to 80%) from total pixels respectively. By this analyzed algorithm, the binary watermark image is extracted from random permutations Fetal ECG audio file has PSNR =15 dB or less. Therefore it gives the analyzed algorithm usefulness in a wide application for contours where the telemedicine is necessary. The higher compression in the first stage is achieved and obtained without any significant losses of the approximation quality.

References

[1] Lu, Z.; Kim, D.Y.; and Pearman, A. (2000). Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. *IEEE Transactions on Biomedical Engineering*, 47(7), 849-856.

[2] Datar, S.; Datar, A.; and Jain, A. (2013). Performance evaluation of high SLFOR fixed windows in ECG signal compression using optimized cosine modulated filter bank. *Proceedings of International Conference on Innovative Trends in Electronics Communication and Applications*, Pondicherry, India, 128-135.

[3] Deepu, C.J.; and Lian, Y. (2015). A joint QRS detection and data compression scheme for wearable sensors. *IEEE Transactions on Biomedical Engineering*, 62(1), 165-175.

[4] M. Alsenwi, M. Saeed, T. Ismail, H. Mostafa, and S. Gabran, "Hybrid compression technique with data segmentation for electroencephalography data," in 2017 29th International Conference on Microelectronics (ICM). IEEE, dec 2017.

[5] WHO. Statistics On Cardiovascular Disease. Available online: https://www.who.int/en/news-room/fact-sheets/detail/ cardiovascular-diseases-(CVDs) (accessed on 10 January 2021).

[6] Castillo, M.; Conte, B.; Hinkes, S.; Mathew, M.; Na, C.J.; Norindr, A.; Serota, D.P.; Forrest, D.W.; Deshpande, A.R.; Bartholomew, T.S.; et al. Implementation of a medical student-run telemedicine program for medications for opioid use disorder during the COVID-19 pandemic. Harm Reduct. J. 2020, 17, 6.

[7] Mansouri, F. Role of Telemedicine and Telegenetics Framework for the Management of Cancer Patients During the COVID-19 Pandemic. Biointerface Res. Appl. Chem. 2021, 11, 8773–8779. [CrossRef]

[8] A.Peterkova & M. Stremy. (2015). The raw ECG signal processing and the detection of QRS complex. *2015 IEE European Modelling Symposium*, pp. 80-85.

[9] W. S. Besbas, "Contour Extraction, Processing, and Recognition," Ph.D thesis, Poznan University of Technology, Poznan,1998.

[10] A. Dziech, "Contour Extraction and Compression-Selected Topics," in Mobile Robots Perception & Navigation, S. Kolski, Ed. Mammendorf: Advanced Robotic Systems International and pro literatur Verlag, 2007, pp. 187-212

[11] Ali Ukasha, Ali Alshanokie, Alwaled Alsharef and Saleh Abuazoum, ''A Review: Improvement of Segment Distance Ratio Method for Arabic Contour Data Approximation''. IEEE 4th International Conference of Saudi Computers Colleges (4th NCCC), 27-28 March 2021, Taif, Kingdom of Saudi Arabia.

[12] G. Xuan, J.Chen, J.Zhu, Y.Q. Shi, Z.Ni, and W. Su, "Lossless Data Hiding Based on Integer Wavelet Transform," IEEE, vol.1, no.2, pp. 7803-7714, 2002.

[13] G. Jeon, "Watermarking on Bit Plane Arranged Images, "International Journal of Control and Automation, vol. 6, no. 2, April, 2013, pp. 335-

346. G. Jeon, "Watermarking on Bit Plane Arranged Images, "International Journal of Control and Automation, vol. 6, no. 2, April, 2013, pp. 335-346.

[14] C. Obimbo and B. Salami, "Using Digital Watermarking for Copyright Protection," in *Watermarking - Volume 2*, M. D. Gupta, Ed. InTech, 2012, pp. 137 -158.

[15] S. Lee, J. Kim and M. Lee. “A Real-Time ECG Data Compression and Transmission Algorithm for an e-Health Device”. IEEE Transactions on Biomedical Engineering. Vol. 58, Issue 9, pp. 2448-2455. September, 2011. ISSN 0018-9294. DOI: 10.1109/ TBME.2011.2156794

[16] L. J. Hadjileontiadis, “Biosignals and compression standards,” in MHealth, PP. 277–292. Springer, 2006.

[17] R. C. Gonzalez, Digital Image Processing, Second Edition, Addison Wesley, 1987.

[18] A. Asokan, J. Anitha, M. Ciobanu, A. Gabor, A. Naaji and D. J. Hemanth. “Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification: An Overview.” Applied Sciences. Vol. 10, Issue 4207, pp. 1-21. JUN. 2020.

[19] I. Panessai, A. Abdulba. “An Efficient Method of EEG Signal Compression & Transmission Based Telemedicine” Journal of Theoretical and Applied Information Technology. . Vol. 97. No 4, pp. 1060-1070. FEB. 2019.

[20] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 2nd ed. Upper Saddle River, New Jersey: Prentice Hall, 2002.

Impact of Neural Machine Translation on Monolingual and Multilingual Arabic Sentiment Analysis

Amna Elhawil, Suher ElBasha and Nabil Drawil
Department of Computer Engineering
Faculty of Engineering
University of Tripoli
Tripoli, Libya
A.elhawil@uot.edu.ly , S.elbasha@uot.edu.ly, N.drawil@uot.edu.ly

***Abstract*—The performance of monolingual sentiment analysis of Arabic language so far is limited due to the language complexity. The recent researches proposed an alternative method based on translating the Arabic text to English in order to take advantage of the high accuracy of the English sentiment analysis. However, more research is still required to optimize the translation process and analyze it efficiently. This paper demonstrates deeply the monolingual and multilingual Arabic sentiment analysis using two different types of translators: Google translator API and the sequence-to-sequence transformer model. We provide a comprehensive study about the translation process and its challenges. The sentiment analysis uses a large dialectic Arabic dataset from different Arabic regions. The accuracy of the machine learning support vector machine (SVM) algorithm of the monolingual and multilingual Arabic sentiment is compared and discussed.**

Keywords—SVM, Google Translator, MarianMT transformer, Arabic sentiment analysis.

I. INTRODUCTION

Sentiment analysis is a process of identifying the classification of a text. The text could be user reviews or opinions about a product or a subject. The classification could be binary (positive or negative) or multi-class of three or more classes. Normally, the main standard steps of the sentiment analysis are: pre-processing the data, feature extraction and classification step. In order to extract the features, language-aware analysis has to be performed. The studies have shown that the accuracy of English sentiment analysis is very good compared with other languages [1] and [2]. However, few studies have dealt with Arabic sentiment analysis [3]-[9]. The achieved accuracy of processing Arabic text with the standard sentiment analysis steps was very limited. The main challenge facing Arabic sentiment analysis is the nature of the Arabic language. Unlike English language, the Arabic language has unique features such as:

1. There are two basic types of Arabic language: Modern Standard Arabic (MSA) and Dialectical Arabic (DA). MSA is the official language whereas DA is the informal language used in the speaking and social media. Each type has its own structure.
2. Arabic language does not have upper and lower cases, so it is hard to distinguish names from it.
3. A single word can have multiple meanings, as example the word 'شعر' means poetry, feeling or hair.
4. In English language there are some words that refer to the same meaning, but they have different letters such as 'Book', 'Library', and 'Write'. Whereas in Arabic language the same words that have same meaning may share a portion of their letters such as,' كتب' means to write, 'مكتبة' means library, and 'كتاب' means book [4].

In our previous paper [10], we presented the limitations of the Arabic sentiment analysis. We reported that Arabic sentiment analysis needs to be enhanced and improved. However, another track has been recently suggested by many studies [6], [7] and [11]. The basic idea concerns translation of the Arabic text to English, then processing it as English text. However, two types of sentiment analysis are introduced: monolingual and multilingual sentiment analysis. In monolingual sentiment analysis the classifier algorithm is trained and tested on a monolingual corpus [12]. On the other hand, in multilingual sentiment analysis the classifier could be trained on a large corpora of a specific language (normally English language); then different language corpora is translated to the language of the trained model and used for testing. The classifier does not have to recognize the language of the original text [12] and [13]. For both types of sentiment analysis, the performance has been enhanced and the classifiers are able to capture the class of the translated text. In addition, the observed results showed that the sentiment analysis of English translated dataset gives competitive results with respect to that of Arabic dataset [7]. Of course, the used translators have a big impact on the performance of the sentiment analysis. For some languages, the quality of the translators increases the performance of the classification [14].

In this paper, we evaluate the performance of the Arabic monolingual and compare it with multilingual sentiment analysis that uses two types of translators: Google translator API and Marian model. Support vector machine (SVM) algorithm that relies entirely on *n*-gram features is used. The remainder of the paper is organized as follows: Section 2 describes in details the neural machine translation development. Section 3. explains the sentiment analysis process. Section 4 presents and discusses the results. Finally, Section 5 draws the conclusions.

II. NEURAL MACHINE TRANSLATION

Google's phrase-Based machine translation (PBMT) is the first translator developed by Google in 2000. It is basically a word-to-word or phrase-to-phrase mapping translator. Although it works well for simple sentences, It's based rule is building dictionaries its translation quality drops down with complex sentences. In 2010, neural machine translation models (NMT) were invented. NMT considers two main transfer approaches: syntactic transfer and semantic transfer. In the syntactic transfer, the input sentence is analyzed to parse trees based on the language grammar rules. In the semantic transfer the common meanings and dependencies are included. The main advantage of NMT is that, it is language independent because the syntactic and semantic analysis can

be reused in translation into other languages [15]. It relies on deep learning where a large amount of training data is used to establish a mapping relationship between the source language and the target language [16]. NMT is called *Encoder-Decoder* architecture. Both the encoder and decoder are composed of a stack of several identical layers. The encoder network encodes the words of the source sentences to a list of vectors. The vectors represent the meaning of the words. On the other hand, the decoder converts back the vectors to words of the target language [17].

In order to capture the time dependence in a sentence, Recurrent Neural Network (RNN) is used. RNN nets in the hidden layer feed back the output of time t to the next input at time *t*+1, as shown in Fig. 1. That tells the network what happened before along with what is happening "now" [18]. Furthermore, Long Short-Term Memory (LSTM) has successfully improved long-sentence translation quality [19]. Nevertheless, to gain significant advantages, the bidirectional neural network (BNN) was proposed in 2015 [20]. BNN translation modeling processes information in backward and forward directions; thus both the past and the future contexts are considered in translations [20].

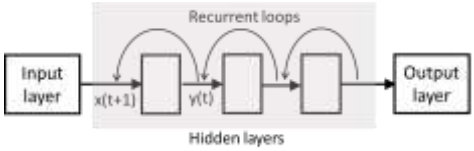


Fig. 1. Recurrent Neural Network

Moreover, attention mechanisms have been introduced by [21] to improve the performance of the translation. An attention layer(s) is added to the decoder as a hidden layer, as illustrated in Fig. 2. This layer allows the decoder to observe other words in the input sentence as it encodes a specific word. This makes the focus on the parts of the sentence that contain the important information. Here, the decoder decodes the input sentences to vectors and weights provided by the attention-mechanism. However, there are two types of models: sequence-to-sequence architecture and transformers. The architectures that use decoder-encoder with or without attention mechanisms are called *sequence-to-sequence* models. The attention architectures are called *transformers*.

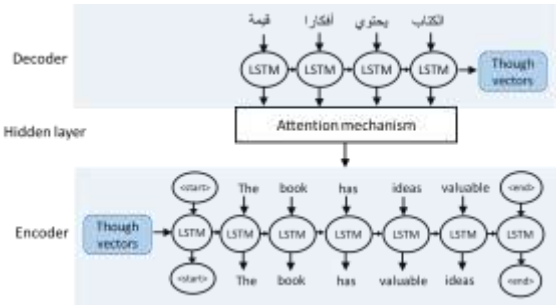


Fig. 2. Attention mechanism layer

Compared to RNNs, the transformer model has proven to be superior in quality for many sequence-to-sequence tasks while being more parallelizable [22]. There are different translators that are available for online and offline use. Google Cloud Platform (GCP), Amazon (AWS) and Microsoft Azure translators are the most common deep learning models. They are designed for online use. These commercial providers have been proven to give consistent translation quality for Arabic, Mandarin Chinese, Persian, and Russian languages [23]. For offline environment, there are different open source NMT models such as: Google Translator , Hugging Face, MarianMT [24], OpenNMT and Fairseq transformers.

A. Google Translator model

Google started with Phrase-Based Machine Translation (PBMT). PBMT breaks an input sentence into words and phrases to be translated largely independently. In 2016 Google developed Google Neural Machine Translation (GNMT) [25]. It can identify over a hundred languages and translate words, sentences or webpages. Fig. 3 shows the basic Google translator architecture. It consists of encoder and decoder networks and an attention module between them. Both the encoder and decoder have 8 LTSM layers. The decoder is a combination of an RNN network and a softmax layer. The model replica is partitioned 8-ways and is placed on 8 different GPUs to speed up the training process. The bottom decoder layer output is only for obtaining recurrent attention context, which is sent directly to all the remaining decoder layers. The softmax layer is also partitioned and placed on multiple GPUs [25]. The translation errors are reduced by an average of 60% compared to Google's phrase-based production system [26]. In this work, the unlimited Google Translator model is used. This GNMT is an auto-language detection, hence the language of the input text needs not to be specified.

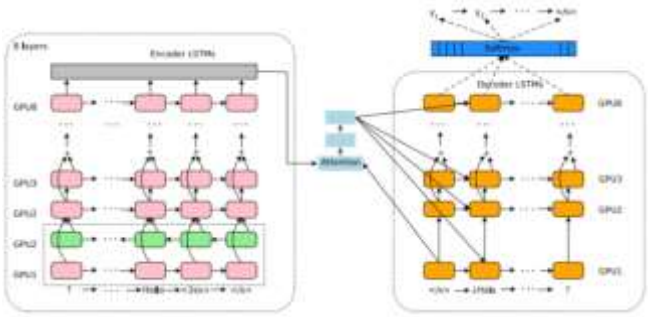


Fig. 3. The model architecture of GNMT system. From [25]

B. Sequence-to-Sequence Transformer model

Sequence-to-sequence transformer models are models that combine RNN and attention as shown in Fig. 4. The performance is enhanced by replacing the LSTMs by full attention structures [27]. In general, there are four types of sequence-to-sequence transformer models:

1. Generic Encoder-Decoder model: The encoder is any pretrained autoencoding model, such as BERT, and the decoder is any pretrained autoregressive model as BERT, ROBERTA, DISTILBERT, CAMEMBERT and ELECTRA [28].
2. MBART model: This model is proposed by [29]. It is multilingual encoder-decoder model pre-trained once on large-scale monolingual corpora in different languages.
3. Marian model: The Marian framework is being developed by the Microsoft translator team. It is transformer encoder-decoder with 6 layers in each component [30].

4. BART model: It is presented by [31] and trained on OPUS dataset [32]. Moreover, in the third analysis

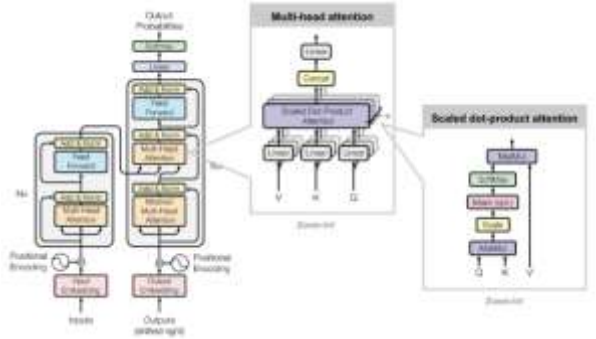


Fig. 4. The model architecture of Transformer, image source [21]

III. SENTIMENT ANALYSIS PROCESS

The Arabic dataset used in this work is a large dataset extracted from 100k Reviews [33]. It includes both modern standard and dialectical Arabic reviews. On the other hand, the English dataset is a set of English tweets and reviews extracted from [34]. Both the Arabic and English datasets are about the same topics which are reviews about hotels, books and movies. The size of both corpus are listed in Table I.

TABLE I. DESCRIPTION OF CLASSES

Corpus name	# Reviews	# Positive classes	# Negative classes
Arabic dataset	6664	3332	3332
English dataset	10000	4903	5097

In general, the sentiment analysis has four basic steps: pre-processing datasets, feature extraction and classification steps. The pre-processing step starts with splitting each sentence into words or tokens, then removing redundancies, emoticons, special characters, symbols, etc. The tokens are converted to numerical feature vectors. The feature extraction technique which is used in this study is *n*-grams technique. In this technique, the sequence of text is broken into parts called *n*-grams. These parts could be single words called 1-gram, two words called 2-gram or three words called a 3-gram. Although 1-gram tokenizer is much faster and efficient, but it does not capture all the information. Higher-grams value keeps more information. The last step of the sentiment analysis is the classification. Support Vector Machine (SVM) algorithm is used for polarity classification because it works extremely well with Arabic sentiment analysis [3] and [8]. In this work, linear kernel is used because it gives higher accuracy than polynomial and sigmoid kernels. All results presented in this article are averages of multiple runs. In monolingual sentiment analysis, the SVM algorithm is trained and tested on a one-language corpora as shown in Fig. 5.

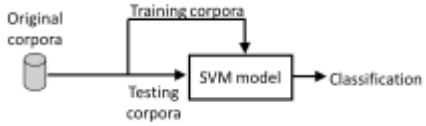


Fig. 5. Block diagram of the monolingual sentiment analysis

In multilingual sentiment analysis, the SVM classifier is trained on English-origin corpus then it is tested using the translated dataset as shown in Fig. 6.

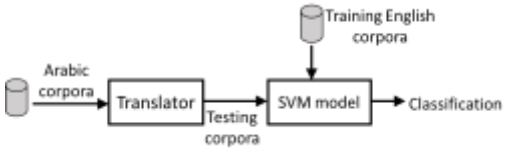


Fig. 6. Block diagram of the multilingual sentiment analysis

IV. RESULTS AND DISCUSSION

As mentioned above, two basic types of analysis are performed in this study: monolingual and multilingual sentiment analysis. The experiments are listed in Table II. In this section we illustrate each experiment and present the corresponding results.

TABLE II. LIST OF EXPERIMENTS PERFORMED IN THIS WORK

Type	#	Training dataset	Testing dataset
Monolingual Sentiment analysis	1	Arabic dataset	Arabic dataset
	2	English dataset	English dataset
	3	Translated Arabic dataset	Translated Arabic dataset
Multilingual Sentiment Analysis	4	Translated Arabic dataset	English dataset
	5	Arabic & English dataset	Arabic & English dataset

A. Monolingual sentiment analysis

Three monolingual sentiment analyses have been performed. The first is done on the original Arabic dataset, where the training and testing datasets are in Arabic language. The second sentiment analysis is performed on the original English dataset. The results of both analyses are shown in Table III. The average accuracies of Arabic and English sentiment analysis are about 80% and 87% respectively.

TABLE III. THE ACCURACY RESULTS OF THE MONOLINGUAL SENTIMENT ANALYSIS

Training Corpora	Testing Corpora	1-gram	2-gram	3-gram
Arabic	Arabic	82%	81%	79%
English	English	87%	88%	87%

Moreover, in the third analysis the Arabic dataset is translated to English. The training and testing datasets are in English. Of course, there are two English versions of the translated Arabic dataset: one is translated using Google Translator and the second is translated using MarianMT transformer. The results of these translated corpus are listed in Table IV. It can be seen that the classification accuracy of SVM for the dataset translated by both translators is mostly consistent. On other words, the translation quality of both translators tends to be equal. Nevertheless, the achieved accuracy for the translated datasets is close to the accuracy of the origin Arabic sentiment analysis. We could say that the information necessary for sentiment analysis is preserved by neural machine translation. This conclusion agrees with that obtained by E. F. Can et al [13] and Rushdi-Saleh et al [35].

TABLE IV. THE SENTIMENT ANALYSIS ACCURACY OF THE TRANSLATED ARABIC DATASET

	Translator	
Feature	**Google Translator**	**MarianMT Transformer**
1-gram	80%	81%
2-gram	81%	79%
3-gram	79%	77%

The percentage of the mispredicted reviews for the three models is about 20%. Table V lists the number of mispredictions of each model. We have studied these reviews for each model and compared between them in order to investigate whether there are common reviews.

TABLE V. THE SIZE OF TEST DATASET AND NUMBER OF MISPREDICTIONS FOR EACH MODEL

Models	Size of test dataset	Number of Mispredictions
Arabic Model	1333	238
Google translate Model	1020	207
MarianMT Model	1333	249

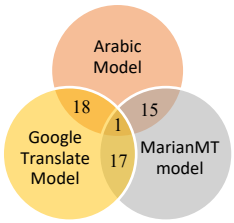


Fig. 7 Number of common mispredictions

The number of common mispredictions for the three models is depicted in Fig. 7. The three models shared around 8% of the mispredicted reviews, and just one review could not be predicted by all three models. The three datasets are individually applied to the SVM classifier in order to improve these results, and the majority forecast of the three models is taken into account. As demonstrated in Table VI, the performance has improved by 2%.

TABLE VI. THE ACCURACY OF THE MAJORITY PREDICTION OF THE THREE MODELS

Feature	Accuracy
1-gram	82%
2-gram	82%
3-gram	83%

In fact, the translation process is expected to enhance the performance of the sentiment analysis to match that of the English sentiment analysis (which is 87%). However, it seems to be some sentences have lost some of their main meaning after the translations. Here are some reasons:

1. Some Arabic words do not have an equivalent English meaning. In such cases Google Translate tends to rewrite the original word with the target language letters [35]. This effects the classification process. Example is shown in Table VII. The dialect word "ديما" (pronunciation: [Dimā]) which means "always ", is translated by Google translator as "Dema". However, the translation of texts in Arabic dialects depends on a good understanding of the text and being aware of the context of the situation [35].

TABLE VII. TRANSLATION EXAMPLE 1

	Sentence
Arabic review	الموقع جدا ممتاز لولا عوامل ممتاز وانشاء الله وفندقي ديما
Human translation	Exceptional. The location is very excellent and excellent for families, and God willing, it will be always my hotel.
Google Translator	Exceptional. The location is very excellent and for excellent and God willing, my hotel is Dima
MarianMT Transformer	The site is very special. The families are excellent. God willing, it is my hotel.

2. Some positive text may be translated from Arabic to English in the form of negation without the intent being the negation in itself. As an example, see the Arabic sentence shown in Table VIII. The translation of Google Translator is classified as negative while it should be positive because the translation gives negative meaning by the word "no thing" which means "is not a thing on anywhere".

TABLE VIII. TRANSLATION EXAMPLE 2

Arabic review	مميز. قريبه من الروضه لاشيء. لاشيء
Human translation	Special. His proximity to the Holy room. Nothing
Google Translator	Special. Close to the honorable. No thing
MarianMT Transformer	Special... near the honorable kindergarten ... nothing

3. There are some Arabic words that have more than one meaning and may not have the same impactful repercussions in English. An example is shown in Table IX. Although this review is negative, but MarianMT model translation gives a positive meaning.

TABLE IX. TRANSLATION EXAMPLE 3

Arabic review	لم يكن على مستوى الطموح. القرب من الحرم وافطار. النظافة وخدمة الغرف
Human translation	It was not up to expectation. proximity to the sanctuary and breakfast. Cleanliness and room service
Google Translator	It was not at the expectation level. The proximity of the Haram and the breakfast. Hygiene and room service.
MarianMT Transformer	It was not as predictable as being close to campus and breakfast, cleanliness and room service.

4. Furthermore, there are some Arabic sentences contain rectification that includes the opposite meaning of the review. Table X shows an example. The review is originally negative. The SVM algorithm classified the translation of Google translator correctly as negative. In contrast, since the MarianMT transformer translation has a positive meaning so it is classified wrongly as positive.

Additionally, it is critical to emphasize that people like to convey their feelings using emoticons. Emoticons appear in about 62 percent of the Arabic tweets used in this study. As a result of the emoticons being deleted during the pre-processing step, the tweets have lost some of their meaning and may be categorised wrongly. One method recommended by [9] to prevent this problem is to translate all emoticons to the words that correspond to them [9].

TABLE X. TRANSLATION EXAMPLE 3

Arabic review	نتمنى الاهتمام بالسياحة الداخلية. الهدوء. المنتجع يحتاج للصيانة نفس مشاكل الصيانة تتكرر في الشاليهات. كما نحتاج فعاليات تخص النساء مثلا مسبح خاص مغلق يحترم خصوصية تقاليد المجتمع.
Human translation	I wish to pay attention to internal tourism. Calm. The resort needs maintenance. The same maintenance problems are always repeated for the chalets. We also need activities for women, for example, a private closed swimming pool that respects the privacy of the traditions of society.
Google Translator	I wish to pay attention to internal tourism. Calm. The resort needs to be repeated as the same maintenance problems for your chalets, we need activities for women, for example, a closed private swimming pool respecting the privacy of community traditions.
MarianMT Transformer	I hope to take care of the interior tourism, the quiet, the resort needs maintenance, the same maintenance problems as your chalets, we need women's activities, such as a closed private pool that respects the privacy of society's traditions.

B. Multilingual Sentiment Analysis

In the multilingual sentiment analysis, two experiments are performed:

1. In the first experiment, the system is trained using the English dataset. It consists of 8000 tweets. This is to make sure the SVM algorithm is trained on large corpus. The tested data is the translated Arabic dataset using Google translator and MarianMT transformer.
2. The original Arabic and translated datasets are combined in the second experiment, resulting in a multilingual corpus. The total number of tweets/reviews in this dataset is 13200, with 3300 positive and 3300 negative tweets/reviews in each language. It is evenly distributed between positive and negative Arabic and English tweets. The classifier is trained and tested using this corpus.

Table XI shows the results of the SVM algorithm for the first experiment. As can be shown, when compared to the monolingual sentiment analysis, the performance lowers to around 72 percent on average. Despite the fact that the SVM method was trained on a huge corpus, it was unable to predict the class of around 28% of the dataset evaluated. The classifier is trained on an English dataset and evaluated on a separate culture-based corpus in this experiment. There are also distinctions in the way words are constructed and in grammar. We hypothesize that the decline in accuracy is related to the linguistic and cultural differences between Arabic and English, as discussed in the previous section.

TABLE XI. THE ACCURACY RESULTS OF THE MULTILINGUAL SENTIMENT ANALYSIS

Training Corpora	Testing Corpora	Translator	1-gram	2-gram	3-gram
English dataset	Translated Arabic	Google Translator	72%	71%	70%
English dataset	Translated Arabic	MarianMT Transformer	72%	72%	71%

In the last experiment, the English and Arabic datasets are merged. The main purpose of this analysis is to test the language detection capability of the SVM algorithm. the size of dataset is listed in Table XII. It is balanced size dataset, consists of 6600 English tweets/reviews with 50% positive and 50% negative tweets and the same number for translated Arabic tweets/reviews. The results are shown in Table XIII. It is clear that, the performance is improved (about 84%) compared with the previous experiment (which is about 70%). In other words, mixing the Arabic and English datasets gives an accuracy of about 84%. That means the classifier was able to detect the language of the tweet/review and predict the class successfully.

All the achieved results are summarized in Table XIV. By comparing the results achieved by [13] for Spanish-English multilingual sentiment analysis it can be noticed that the obtained average accuracy in [13] was about 60% although the high presence of English words in Spanish language. In this work, the accuracy is higher, although there is a big difference between Arabic and English words. Moreover, the translation process of Arabic language to English causes the accuracy to drop by 4% in [35] which in this work the drop is only 1%. However, the results showed that it is possible to avoid the translation mistakes and problems by mixing different language datasets, and the SVM classifier has the ability to detect the language and make its predictions successfully.

TABLE XII. DESCRIPTION OF MIXED DATASET

Training dataset	Testing dataset	Positive tweets	Negative tweets
10560	2640	6600	6600

TABLE XIII. THE SENTIMENT ANALYSIS ACCURACY OF THE MIXED CORPUS

Feature	Accuracy
1-gram	83%
2-gram	85%
3-gram	84%

TABLE XIV. SUMMARY OF THE OBTAINED RESULTS

Type	#	Training dataset	Testing dataset	Average accuracy
Monolingual Sentiment analysis	1	Arabic dataset	Arabic dataset	81%
	2	English dataset	English dataset	87%
	3	Translated Arabic dataset	Translated Arabic dataset	80%
Multilingual Sentiment Analysis	4	Translated Arabic dataset	English dataset	71%
	5	Merged Arabic & English dataset	Merged Arabic & English dataset	84%

V. CONCLUSION

In conclusion, we looked into the impact of two translation models on Arabic sentiment analysis: Google translator and MarianMT transformer. The results revealed that translation had no substantial impact on the classifier's performance. Furthermore, the SVM classifier trained on the original English dataset could not perform well with the translated non-original English data for multilingual sentiment analysis. SVM was also able to detect the language of the mixed language dataset, which was a promising finding. The integrated English-Arabic dataset achieves the best results, with an average of 84 percent. SVM, on the other hand, can successfully classify tweets/reviews in several languages without the requirement for translation. However, it may be

interesting to perform the same study using other classifiers such as Naïve Bayes (NB), Long Short Term Memory (LSTM), and BERT.

REFERENCES

[1] A. Nwesri, S.M.M. Tahaghoghi and F. Scholer, "Stemming Arabic Conjunctions and Prepositions," in Proceedings of the 12th international conference on String Processing and Information Retrieval, Heidelberg, pp. 206-217, 2005.

[2] R. M. Elawady, S. Barakat, H. M. El-Bakry, and M. Elrashidy, "Sentiment Analysis for Arabic and English Datasets", International Journal of Intelligent Computing and Information Science (IJICIS), vol.15, No. 1, January 2015.

[3] O. Oueslati, E. Cambria, M. B. HajHmidaa, and H. Ounellia, "A Review of Sentiment Analysis Research in Arabic Language", Future Generation Computer Systems, vol. 112, pp. 408-430, November 2020.

[4] Z. Aizouky, "Arabic-English Google Translation Evaluation and Arabic Sentiment Analysis", Master of Arts, Department of Digital Humanities, University of Alberta, 2020

[5] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic Sentiment Analysis: A Systematic Literature Review", Applied Computational Intelligence and Soft Computing, vol. 2020, pp. 1-21, January 2020.

[6] M. Salameh, S. M. Mohammad and S. Kiritchenko, "Sentiment after translation: A case study on Arabic social media posts", In: Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies. pp. 767–777, 2015.

[7] A. Barhoumi, C. Aloulou, N. Camelin, Y. Esteve and L. H. Belguith, "Arabic Sentiment analysis: an empirical study of machine translation's impact", LPKM, 2018.

[8] Anwar Alnawas, "The Corpus Based Approach to Sentiment Analysis in Modern Standard Arabic and Arabic Dialects: A Literature Review", Journal of Polytechnic, vol. 21, Issue 2, pp. 461-470, 2018.

[9] M. A. Sghaier and M. Zrigui, "Sentiment analysis for Arabic e-commerce websites," 2016 International Conference on Engineering & MIS (ICEMIS), 2016, pp. 1-7, doi: 10.1109/ICEMIS.2016.7745323.

[10] A. Elhawil, Y. Trabelsi and M. Mahfoud, "Comparison between the NB and SVM Methods for Multiclass Arabic Sentiment Analysis," 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 2021, pp. 913-917, doi: 10.1109/MI-STA52233.2021.9464487.

[11] S. M. Mohammad, M. Salameh and S.Kiritchenko, "How translation alters sentiment", Journal of Artificial Intelligence Research, vol. 55, pp. 95–130, 2016.

[12] D. Vilares, M. A. Alonso and C. Gomez-Rodrıguez, "Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora", 6th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis. WASSA 2015. Workshop Proceedings, 2015.

[13] E. F. Can, A. Ezen-Can and F. Can, "Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data", arXiv:1806.04511, 2018.

[14] A. Balahur and M. Turchi, "Comparative Experiments for Multilingual Sentiment Analysis Using Machine Translation", CEUR Workshop Proceedings, vol. 917, pp. 75-86. JRC73295, 2012.

[15] N. Klyueva, "Semantics in Machine Translation", WDS'07 Proceedings of Contributed Papers, Part I, pp. 141–144, 2007.

[16] Tencent Research Institute, CAICT, Tencent AI Lab and Tencent open platform, *Artificial Intelligence: A National Strategic Initiative,* pp. 40, 1st ed. 2021.

[17] Y. Wu, M. Schuster, Z. Chen, Q. V Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", arXiv: 1609.08144v2, 2016.

[18] H. Lane, C. Howard and H. M. Hapke, *Natural Language Processing in Action Understanding, analyzing, and generating text with Python*, 1st edition, pp. 253, 2019.

[19] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions", European Journal of Operational Research, vol. 270, issue 2, pp. 654–669, 2018.

[20] A. Peris and F. Casacuberta, "A Bidirectional Recurrent Neural Language Model for Machine Translation", Procesamiento de Lenguaje Natural (SEPLN) 2015, 2015.

[21] Ashish Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need", 31st Conference on Neural Information Processing Systems (NIPS 2017), USA, 2017, arXiv:1706.03762

[22] F. Chollet, "Deep Learning with Python", 2ed Edition, pp. 358, 2021.

[23] A. Ai, "Evaluating Machine Translation Providers",2020, accessed 4 Dec. 2021, <https://anno-ai.medium.com/evaluating-machine-translation-providers-68b8430debc3>.

[24] M. Junczys-Dowmunt, R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. F. Aji, N. Bogoychev, A. F. T. Martins and A. Birch, "Marian: Fast Neural Machine Translation in C++ ", The 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15-20 July 2018.

[25] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. R. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. S. Corrado, M. Hughes and J. Dean, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, ArXiv, vol. 1609.08144, 2016.

[26] R. Sagar, "Machine Learning Behind Google Translate", 2020, Analytics India magazine, accessed 4 Dec. 2021, <https://analyticsindiamag.com/google-translate-machine-learning/>

[27] Á. Rocha, H. Adeli, G. Dzemyda, F. Moreira, A. M. R. Correia, "Trends and Applications in Information Systems and Technologies", Springer International Publishing, vol. 2, pp. 271, 2021.

[28] "Encoder Decoder Models", 2020, huggingface, accessed 4 Dec. 2021,<https://huggingface.co/transformers/v3.5.1/model_doc/encoderdecoder.html>.

[29] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis and L. Zettlemoyer, "Multilingual Denoising Pre-Training for Neural Machine Translation." Transactions of the Association for Computational Linguistics, vol. 8, 2020, pp. 726–742., https://doi.org/10.1162/tacl_a_00343.

[30] T. Rajapakse, "Seq2seq Model." Simple Transformers", 2020, https://simpletransformers.ai/docs/seq2seq-model/.

[31] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, "Bart: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.

[32] P. Chowrasia, "English to Hindi Text Translation Using MARIANMT Models from Huggingface." Medium, Geek Culture, 20 July 2021, https://medium.com/geekculture/english-to-hindi-text-translation-using-marianmt-models-from-huggingface-187b4719809e.

[33] https://www.kaggle.com/abedkhooli/arabic-100k-reviews?select=ar_reviews_100k.tsv, Retrieved Oct 2021.

[34] https://github.com/Gunjitbedi/Text-Classification/blob/master/corpus.csv, Retrieved Oct. 2021.

[35] M. Rushdi-Saleh, M. T. Martín-Valdivia, L. A. Ureña-López, José M and Perea-Ortega , "Bilingual Experiments with an Arabic-English Corpus for Opinion Mining", Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, 740–745

Face Recognition Performance Analysis Study with Flipped and Symmetrical Samples

Khaled Elgdamsi
University of Tripole
Tripole, Libya
K.elgdamsi@uot.edu.ly

Mustafa Alwefati
University of Tripole
Tripole, Libya
mustafaalwefati66@gmail.com

***Abstract* — Face Recognition is a technology that enables people to identify faces using an image or a video clip. Although it has been around for a long time, the task of developing a robust and reliable system is still a bit challenging, the human face can have only a limited number of training images. In this paper, a new set of training samples is generated from the original samples to increase number of used samples by using mirror and the symmetry property of the face. The Histograms of Oriented Gradients (HOG) and The Local Binary Pattern (LBP) methods are used to extract the features from the face images, in general, increasing the number of training images upturns the performance of face recognition systems and the recognition performance is improved. The proposed methods is tested and evaluated using OLR dataset which is widely used for testing and comparing the accuracy of face recognition systems. The experimental results show that the proposed method has a recognition accuracy rates higher than the traditional methods.**

Keywords—face recognition, machine learning, hog, lbp.

I. INTRODUCTION

Humans are identified by their physiological, behavioral, and biological properties by biometrics. The necessity for biometric authentication is due these difficulties with traditional password systems. Biometric systems are compatible to all applications that need authentication mechanism. Security identification accounts for the demanding to use multibiometric systems, which include several types of recognition, rather than non-biometric systems. Face Recognition (FR) is a very efficient and widely used identification technology which is considered as a reliable and very high accurate biometric system. Facial Recognition is a Biometric Artificial Intelligence based application that can uniquely identify a person by analyzing patterns based on the person's facial textures and shape. A facial recognition system is a technology capable of identifying or verifying a person from a video source. FR is one of the most active research fields of computer vision and pattern recognition, with many practical and commercial applications including identification, access control, forensics, and human-computer interactions. However, identifying a face in a crowd raises serious questions about individual freedoms and poses ethical issues. Significant methods, algorithms, approaches, and databases have been proposed over recent years to study constrained and unconstrained face recognition. 2D approaches reached some degree of maturity and reported very high rates of recognition. This performance is achieved in controlled environments where the acquisition parameters are controlled, such as lighting, angle of view, and distance between the camera–subject. Feature extraction step consists of extracting from the detected face a feature vector named the signature, which must be enough to represent a face. The individuality of the face and the property of distinguishing between two separate persons must be checked [1,2]. In this paper, the training set is increased by using two methods, using the flipped and symmetry face techniques. The Histograms of Oriented Gradients (HOG) and The Local Binary Pattern (LBP) methods are used to extract the features from the face images. There are many methods can be used to extract the features from the face images such as: the Local Binary Pattern (LBP), the Gray Level Co-Occurrence Matrix (GLCM), the Gabor Filter, and the Histograms of Oriented Gradients (HOG), since these methods perform well for a texture feature extraction that could be used for the FR. The performance of FR methods can be tested and examined using some benchmark facial datasets such as: Olivetti Research Laboratory (ORL), GT and AR datasets.

II. METHOD

In order to increase the size of the training data, new training images are generated using the mirror and symmetry property of the face,

A. Mirror Face Property

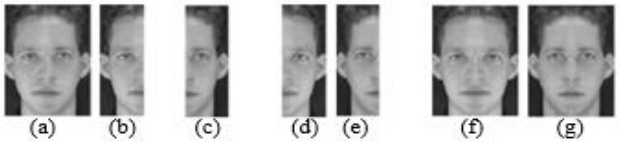
The human face has two sides, right and left, and using the mirror feature, the right sides can be turned to become left and vice versa as shown in figure 1. Using this feature, new images of the same person can be created, which are almost identical to this person.

(a) (b)

Figure 1 : a) original face image, b) flipped face image

B. Symmetry Face Property

Facial symmetry is one specific measure of bodily symmetry. Along with traits such as averageness and youthfulness it influences judgments of aesthetic traits of physical attractiveness since those images reflect some appearance of the face that is not shown by the original images as illustrated in **Figure**.2.



(a) (b) (c) (d) (e) (f) (g)

Figure 2: a) Original image, b) left side, c) right side, d) mirror of left side, e) mirror of right side, f) integrating left side with mirror, g) integrating right side with mirror.

III. Feature Extraction

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing.

A. Local binary pattern (LBP)

One of the most widely used methods to analyze and model texture is the LBP method. It could be basically described as a 3 × 3 square operator. In each square, the eight-neighborhood pixels were compared with the one in the center. If the pixel values of the neighbors were greater than or equal to the pixel value at the center, they were replaced by 1. If not, then their values were replaced by 0. Then, the new binary values of the neighbors were concatenated to produce one decimal value that was considered to be a new value for the pixel in the center. The window was passed to the next pixel and the same operation was repeated. These new decimal values represented the histogram of the input texture [3,4]. the algorithm of the LBP operation is given by:

$$LBP_{N_P,R(x,y)} = \sum_{N_P=0}^{N_P-1} s(g_P - g_C)\, 2^{N_P} \qquad (3.1)$$

where s is the sign function, N_P is the number of neighborhood pixels, g_p represents the gray level value of the neighboring pixels, and g_c represents the gray level value of the central pixels. 2^P is required to produce decimal values.

The traditional LBP analyzes the texture of the image and thresholds a 3 × 3 square neighborhood as the center pixel value. It only uses the sign information to produce the LBP , as illustrated in Figure 3

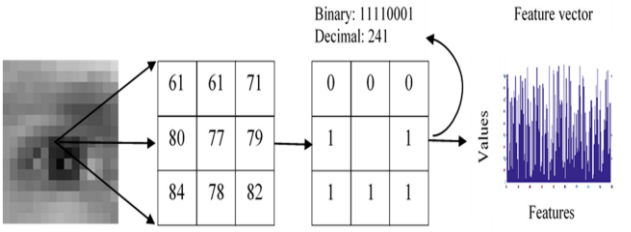


Figure 3: The Local Binary Pattern (LBP) architecture

In a newer implementation, the LBP operation has been upgraded to deal with any neighborhood size, by replacing the square with a circle. This can be described by (NP,R), where In a newer implementation, the LBP operation has been upgraded to deal with any neighborhood size, by replacing the square with a circle. This can be described by (NP,R), where R is the radius of the circle used. Figure 4 illustrates an (8, 2) neighborhood. Additionally, there are a number of other modifications to the LBP.

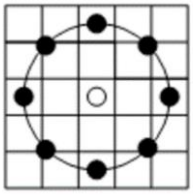


Figure 4: Circular (8, 2) neighborhood

The term *LBP*Np, R is used to describe the LBP operation, the resulting histogram results in the necessary information distributed in the image, such as edges, corners, uniform areas, etc. The effective operation must take care of the spatial information in the image, during the representation. One strategy to accomplish this is to partition the image into a number of small areas *R*0, *R*1, …, *R*m-1, where m is the number of areas. If the size of the histogram is B, then the length of the feature vector is mB. It is obvious from this relation that the number of areas m determines the length of the feature vector, which means selecting small areas results in long feature vectors, leading to extreme use of memory and a slow classification processing. Selecting large areas causes a loss of spatial information. An example of a preprocessed face image partitioned into thirty-six windows and the resulting face feature histogram are illustrated in Figure 5.

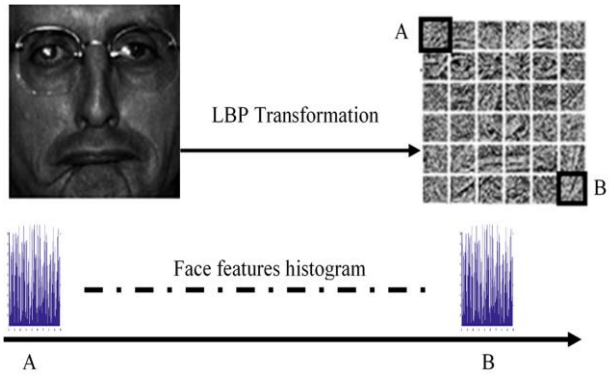


Figure 5: Example of a preprocessed face image partitioned into thirty-six windows and its feature histogram using the Local Binary Pattern (LBP)

B. Histograms of Oriented Gradients (HOG)

One of the very popular methods for feature extrication is the Histograms of Oriented Gradients (HOG). It is one type of descriptors that is used a lot in the human detection.

The HOG concept is to compute the gradient orientation and the gradient direct magnitude. To obtain the HOG of an image, first, the changes in X and Y are computed, then the magnitude and direction are obtained [5].

- Computing Gradients

The main operation of HOG is the derivative, or the center difference, since, there are two derivatives, the x derivative and the y derivative, once these derivatives are obtained, the gradient magnitude and the gradient orientation can be computed.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h} \qquad (3.2)$$

The magnitude is given by:

$$s = \sqrt{s_x^2 + s_y^2} \qquad (3.3)$$

And the orientation is given by:

$$\theta = \arctan\left(\frac{s_y}{s_x}\right) \qquad (3.4)$$

- Blocks and Cells

Figure 6 shows a face image, it is assumed that this image is a 64x128 image, if this image is divided to 128 cells, then some blocks are taken, for example the first block is block 1 with 2x2 cells, then the second block is 50 % overlapped, which block 2, so, each block is consist of 2x2 cells with size 8x8 which means 16x16, with 7x15 = 105 blocks in total.

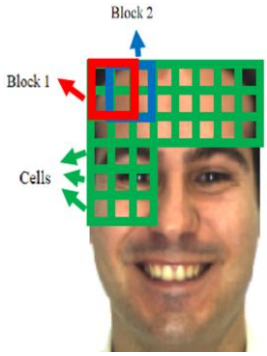


Figure 6: The blocks and the cells

C. HOG Feature Extraction Steps

To calculate the HOG for an image with 64 x 128, for example, the image is divided onto 16x16 blocks with 50% overlap, so therefore there are 7x15 with total of 105 of blocks, and each block consists of 2x2 cells, and the size is 8x8, then the HOG is quantized with 9 directions or bins, if the direction is not one of the bins then some kind of interpolation can be done, also, the Gaussian can be applied to smooth the histogram, then all the descriptors can be concatenated since there are 105 of these block and each one is 9 dimensional, this gives a very large described, about 3780 dimension descriptor and this for the whole image of the block in the image, as shown in Figure 7

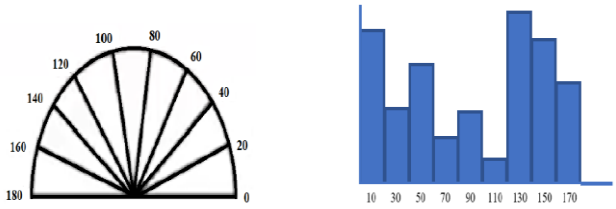


Figure 7: The histogram quantization to 9 bins

The procedure of HOG for feature extraction can be summarized in the following steps:

i. Compute the centered horizontal and vertical gradients with no smoothing.

ii. Compute gradient orientation and magnitudes.

• For color image, pick the color channel with the highest gradient magnitude for each pixel.

iii. For 64x128 image,

iv. Divide the image into 16x16 of 50% overlap.

• 7x15=105 blocks in total

v. Each block should consist of 2x2 size 8x8

vi. Quantize the gradient orientation into 9 bins.

• The vote is the gradient magnitude.

• Interpolate votes bi-linearly between neighboring bin center.

• The vote can also be weighted with Gaussian to down-weight the pixel near the edges of the block.

Concatenate histograms (Feature dimension: 105x4x9 = 3780)

d. The Linear Interpolation

The better histograms can be found by doing the interpolation. If there are 9 bins, and the range of the gradient orientation is between [0º 180 º], this range is quantized into these 9 bins, if the orientation has 85 º, and since, there is no bin with 85º, in this case, this is split into couple of bins which are closest to that, these bins are 70 º and 90 º, since the difference between 70º and 85º is 15 º , and the difference between 90º and 85º is 5 º, the values is divided proportionally according to this ratio, that means (5/20 = 1/4) and (15/20 = 3/4), and the histogram is distributed according to this concept, as shown in Figure 8

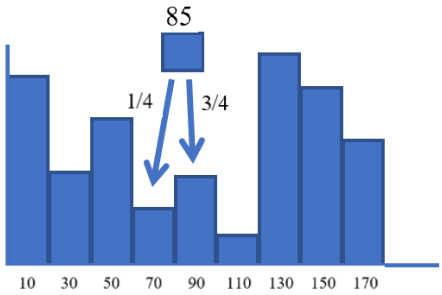


Figure 8: The HOG interpolation

- Feature Vector

Each block has its histogram; all the histograms are concatenated to produce the final feature vector of the whole image as shown in Figure 9

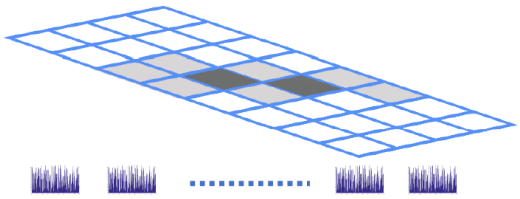


Figure 9: Concatenating the histograms

- Visualization

The visualization of the HOG, as in Figure 10, with some blocks and each block has its histogram which corresponds to the face regions, and some blocks give the dominating direction for some certain region, which give the visualization of the representation that represents the face and calculates how much the distance from these different parts of the face.

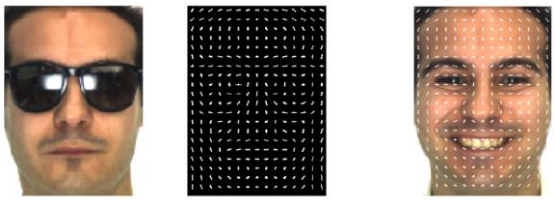


Figure 10: The HOG visualization

By looking to the presentation (middle image), it is clear it is representing a face, and this is used to recognize the face. Once the descriptor exists, for lots of training examples (faces), then any train techniques such of machine learning can be used for FR by classifying the face according to their features.

The HOG is a very strong and popular descriptor, and it is a kind of global descriptor which looks at the whole image. Authors propose this descriptor and they use it for face detection and FR [6].

- Classification

Support Vector Machine is a supervised learning tool commonly used in classification and regression problems. A computer program that uses support vector machines may be asked to classify an input into one of two classes. The program will be provided with training examples of each class that can be represented as mathematical models plotted in a multidimensional space.

The program plots representations of each class in the multi-dimensional space and identifies a hyperplane or boundary which separates each class. When a new input is analyzed, its output will fall on one side of this hyperplane. The side of the hyperplane where the output lies determines which class the input is. This hyperplane is the support vector machine [7,8].

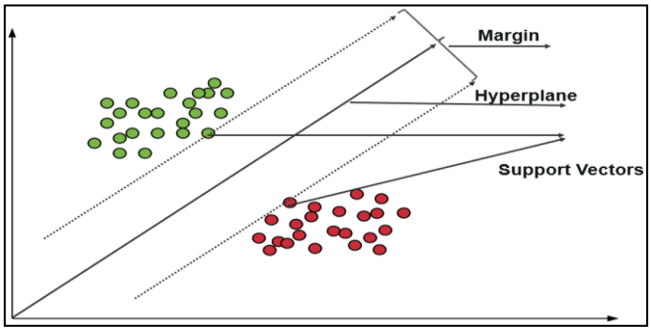


Figure 11 support vector machine

- **Olivetti Research Laboratory ORL**

The ORL is a well-known face dataset that is used to test FR algorithms. It has 400 images of 40 distinct persons, 10 images for each person. The dataset is varied in many aspects. First, the images are taken at different times during the lives of the people. Second, the images include different variations and different facial expressions, such as closed or open eyes, some are smiling, others are not. In addition, there are a number of people wearing spectacles while others are not wearing spectacles. Furthermore, a number of the images include up to twenty degrees of tilting and rotation of the face as shown in figure 12.

Figure 12: Samples from ORL dataset [9]

IV. RESULTS AND DISCUISSON

These results are carried out using the images from ORL dataset, in this part, all ORL dataset is used to test the recognition system, the system is trained using 10% of dataset and the rest 90% is used for testing, then 20% for training and 80% for testing, up to 95% for training and 5% for testing. In this experiment, the Local Binary Pattern (LBP) and Histograms of Oriented Gradients HOG are used for feature extraction, this technique extracts a lot of feature from the image and produce a long feature vector that describe the information in the image. The results are shown in Figure 8. The recognition system is examined one time using original training samples (OTS) and the second time using original with symmetrical training samples (OSTS).

Table1: The Recognition Rate of the ORL dataset using LBP with SVM.

Type of training samples	Training set								
	1	2	3	4	5	6	7	8	9
	Recognition Rate %								
(OTS)	68	76	87	91	92	92	94	94	95
(OFTS)	74	81	90	91	92	93	96	96	96
(OSTS)	73	82	90	92	92	94	96	96	96
(OFSTS)	74	82	90	92	92	94	96	96	96

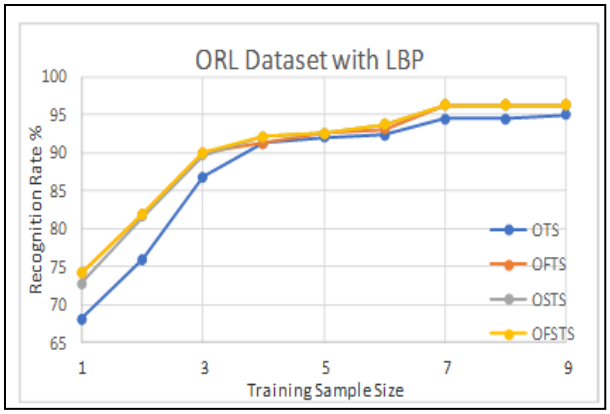


Figure 13: Recognition rate using LBP

As we can see from Figure 13 and Table 1, as the number of training samples increase, the accuracy increases besides. Results using original with flipped samples or original with symmetrical samples are better than using only the original samples, and when using original with flipped and symmetrical are the best.

Table2: The Recognition Rate of the ORL dataset using HOG with SVM.

Type of training samples	Training set								
	1	2	3	4	5	6	7	8	9
	Recognition Rate %								
(OTS)	55	68	81	88	89	92	95	95	95
(OFTS)	64	72	84	89	90	93	95	96	96
(OSTS)	63	73	85	90	91	93	95	97	98
(OFSTS)	64	73	85	90	91	93	95	97	98

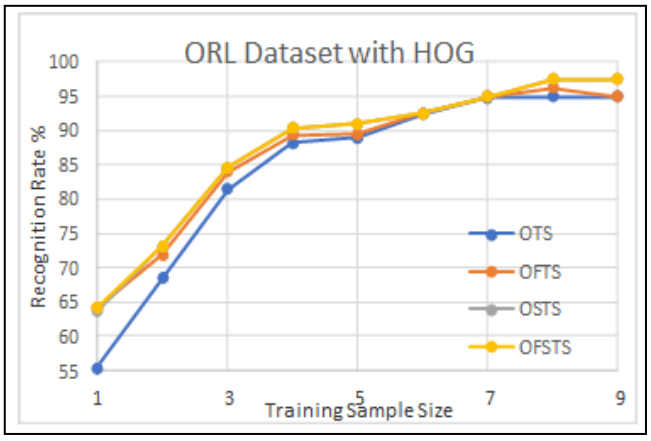


Figure 14: Recognition rate using HOG

As shown in Figure 14 and Table 2, as the number of training samples increase, the accuracy increases. Similarly, the obtained results using original with flipped samples or original with symmetrical samples are better than using only the original samples and when using original with flipped and symmetrical are the best where exceeds 95%.

V. CONCLUSION

This paper presents an effective method to overcome the restricted number of the training sets using the flipped method and the symmetry method.

Increasing the training set sample size with traditional method increase the accuracy, but increasing the training set sample size and using the flipping and symmetry method to generate more train image provide higher accuracy.

REFERENCES

[1] S. Allagwail, Abdulmajed Elbkosh, " Face Recognition with Symmetrical Face Training Samples Based on Histograms of Oriented Gradients" Elmergib University, 2020.

[2] S. Allagwail, O. S. Gedik, and J. Rahebi, "Face Recognition with Symmetrical Face Training Samples Based on Local Binary Patterns and the Gabor Filter" *Symmetry,* vol. 11, p. 157, 2019.

[3] Virgil Petrescu, R. V. " Face Recognition as a Biometric Application" Journal of Mechatronics and Robotics, 3(1), 237-257 (2019)

[4] Waldemar Wójcik, Konrad Gromaszek and Muhtar Junisbekov, "Face Recognition: Issues, Methods and Alternative Applications" 978-953-51-2421-4, 2016.

[5] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," Pattern Recognition, vol. 62, pp. 135-160, 2017.

[6] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," Pattern recognition letters, vol. 32, pp. 1598-1603, 2011.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 886-893.

[8] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in Computer vision-eccv 2004, 2 ed: Springer, 2004, pp. 469-481.

[9] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," Pattern Recognition, vol. 62, pp. 135-160, 2017.

Virtual Reality Lab Technology as a Chemical Engineer Instructional and Mentoring Program Platform at Elmergib University: A Case Study

Mohamed Edali
Department of Chemical and Petroleum Engineering
Elmergib University
Elkhoms, Libya
dredalicreativitygroupts@gmail.com

Asma Milad
Department of Chemical and Petroleum Engineering
Elmergib University
Elkhoms, Libya
asma.aga2009@gmail.com

Zaed Sahem
Ontario Professional Engineering & Creativity Group for Technical Services
Ottawa, Ontario, Canada
sahemzaed@gmail.com

Walid Alaswad
Department of Chemical and Petroleum Engineering
Elmergib University
Elkhoms, Libya
walaswadzo@gmail.com

Ali Baseibsu
Department of Chemical and Petroleum Engineering
Elmergib University
Elkhoms, Libya
ali.bsebsu@gmail.com

Faraj Ben Rajeb
Department of Oil and Gas Engineering
Memorial University
St john's, Newfoundland, Canada
fabr67@mun.ca

Ali Elkamel
Department of Chemical Engineering
University of Waterloo
Waterloo, Ontario, Canada
aelkamel@uwaterloo.ca

***Abstract*—Exploring innovative and cost-effective creative teaching methods is at the forefront of academic research in chemical engineering. Due to the limited physical resources available to engineering students in Libyan universities, sophisticated training techniques are becoming more desirable for successful operator training. Furthermore, events of unforeseen circumstances, the COVID pandemic, for example, necessitates improved educational adaptation and distance learning options. Three-dimensional interactive virtual settings, similar to those used in the videogame and movie industries, are being considered in this aspect. On the other hand, Undergraduate students frequently lack industry expertise, and they are unable to understand the complexity of real-world process units, which may be efficiently used to create learning environments. As a result, students may graduate with insufficient professional hands-on experience. Virtual reality proves to be a revolutionary technology that can satisfy the demands of both academics and industry in this respect. This article concentrates on chemical lab teaching technology pedagogy and introduces virtual reality into Elmergib university's chemical engineering education opportunities and challenges. In order to implement sophisticated immersive learning applications, the paper emphasizes the importance of integrating virtual reality interfaces with computer simulations. In addition, the research highlights the necessity for unique educational effect assessment approaches for virtual-reality-based learning evaluation. Finally, a running case study is shown to completely comprehend and incorporate participants towards using virtual reality technology in chemical engineering education, as demonstrated by a lab safety course.**

Keywords—Virtual reality; Learning design; Process engineering; Educational technologies; Immersive learning.

I. INTRODUCTION

Virtual reality (VR) is a 3D, dimensional interactive computer simulation that tries to completely immerse the user in a virtual environment. It has been used in education for the past decade. However, there is no complete appraisal of virtual reality new emerging technologies in the chemical academic engineering background in the literature. In the instance of academia, Patle *et al.* examined the state and research opportunities for virtual reality contributions in chemical engineering education from 2000 to mid-2017 [1]. So far, this is the only literature review that has been done in this field. On the other hand, the authors neglected to address wider technological and pedagogical difficulties, as well as the numerous mathematical models that may be included in Virtual reality-based systems. This article provides an overview of the use of virtual reality (VR) in chemical engineering education, with a focus on implementations in higher education and the process market. This article is addressing both benefits and drawbacks of using virtual reality in education, focusing on the crucial areas of technology and pedagogy. The purpose of this article is to examine current educational benefit evaluation methodologies as well for engineering and technology virtual space. It would also look into the more significant pedagogical implications and research prospects in this sector. In light of this study's general objective, rather than virtual labs, the evaluation will be applied to virtual-reality-based case studies.

II. LITERATURE REVIEW

In academia, several research attempts have been published in the literature to bring genuine process plants to a university setting using virtual reality technology. However, student plant visits and industrial field trips are difficult to schedule regularly due to logistical challenges and associated expenditures. In addition, students are not permitted to operate process equipment in a real-world environment. As a consequence, most biochemical science and engineering students lack the requisite exposure to process theory and opportunities to apply it to real-life situations. Hands-on training simulators are also important in the process sector for the training of qualified operators. Traditional instruction simulators, on the other hand, lack authenticity and do not provide operatives with practical learning expertise in various unforeseen and risky circumstances [2], [3].

III. VIRTUAL REALITY FOR CHEMICAL ENGINEERING ACADEMIA

Three core domains for chemical engineering education have been identified as having promises and shortfalls as a result of the literature review. That encompasses the fields of technologies, pedagogy, and socioeconomics, among others, as seen in Figure 1. Virtual reality is commonly utilized in chemical curriculum and instruction as a visual aid for displaying three-dimensional images of process units in the technical realm. It might, however, be supplemented with a set of sophisticated mathematical models that allow for advanced techniques in an immersive setting. Concurrently, university students' technical ability to accept virtual reality technology in learning is critical [4]. The literature in the field of education reveals two major difficulties.

To begin with, it recognizes the need to evaluate the impact of virtual reality technology on chemical engineering education. Second, there is a range of groups of students with various learning objectives. Finally, the data reveal that the literature promotes the use of virtual reality technology in chemical engineering education, but that adoption is constrained. The

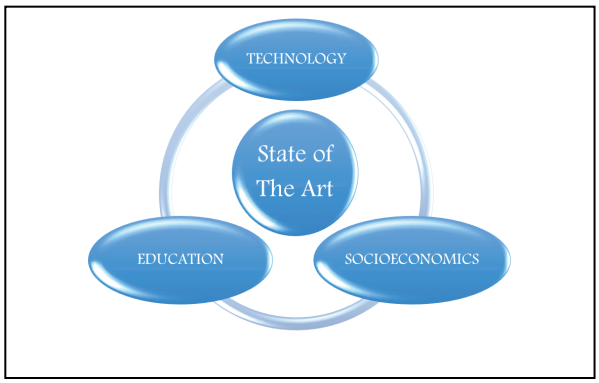


Figure 1: The highlighted domains with research opportunities for chemical engineering education.

challenges that come with them are looked at from the perspective of a broader organizational foundation.

The CAD application is used to generate three-dimensional models of diverse process equipment and operational processes, much like any other computer-aided design; possible technological challenges include developing virtual reality tools for chemical engineering education. Commercial virtual reality software is also used to generate realistic three-dimensional virtual environments using the CAD model as an input. Virtual reality headsets and optical sensors, for example, can help to make the experience even more lifelike. Mathematical models might also be integrated into the virtual reality interface to provide interactive learning activities for educational reasons. The virtual operation of dynamic process plants, the testing of significant mishaps, and the simulation of probable explosions in a chemical reactor are all examples of mathematical models in chemical engineering education (Nortonet al., 2008; Patle et al. 2019). To construct dynamic process simulations and interactive learning modules, users can utilize first-principal models, data-driven models, or a combination of both. Figure 2 depicts a summary of the main modelling methodologies for virtual reality-based systems, as well as their distinguishing characteristics.

First-principle models may be used to represent the genuine behaviour of process plants using comprehensive thermodynamic property models and transport equations. First-principle simulations have been used in research to improve the instructional capabilities of virtual reality-based systems in chemical engineering. Pirola and colleagues used AVEVA's DYNSIM software to create a dynamic simulation of a crude distillation unit facility [5]. DYNSIM is a commonly used programme for using first-principle models to describe the dynamics of process plants. The programme comprises thermodynamic models for the complete process plant, as well as mathematical equations for the many unit processes. In process engineering, data-driven models, also known as machine learning models, are used. A digital twin model is an example of a data-driven model. A virtual version of a physical processor or system that may be used to track it in real-time is known as a digital twin [6]. Hybrid models, in addition to classic first-principle models and data-driven models, must be created to fully capture the complexity and unpredictability of real-world processes [7]. This hybrid method frequently employs a combination of first-principle models and data-driven models to increase the modelling capabilities of immersive process plants. Using virtual reality-based educational systems, a combination of models may be utilized to create a variety of realistic learning scenarios for testing.

IV. EDUCATIONAL VIRTUAL REALITY IMPLICATIONS IN CHEMICAL ENGINEERING

Virtual reality environments in educational contexts are increasingly being used, with potentially significant pedagogical consequences for chemical engineering education. There are two studies, [8] and [9], that suggest potential teaching and learning benefits from the use of such settings. Furthermore,

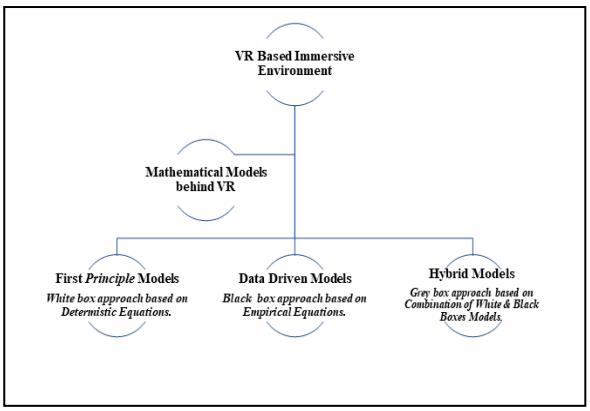


Figure 2: A glance at the various modelling methodologies used in virtual reality systems.

owing to the active rather than the passive way of learning, interactive virtual reality-based settings can provide better benefits [10]. Virtual reality is projected to provide the following notable benefits in the field of chemical engineering education. Students or operators can use the virtual interface to study at their own pace, which is not possible in a real-world process plant and without any space or time limits. The user is allowed to control the virtual camera's movement inside the interactive

learning environment [10]. Virtual reality will enable users to learn in a safe environment while participating in various what-if scenarios that need teamwork and coordination [11], [12]. This assures compliance as well as exposure to a variety of potentially hazardous scenarios. Using a real chemical plant, on the other hand, may not be possible. Process operators may learn a lot by creating virtual avatars or digital representations of themselves to experience teamwork during simulated plant shutdowns, emergencies, and evacuations [13], [14]. Students who participate in multiplayer activities as undergraduates are better equipped to enter the workforce after graduation.

The majority of research focuses on the short-term benefits of virtual reality while neglecting the long-term disadvantages. As a result, innovative approaches for assessing virtual reality's educational impact on a group of users are necessary. The long-term learning benefits and limitations of using immersive environments should be researched further compared to traditional classroom-based education. Students studying chemical engineering at the undergraduate level have a sound scientific foundation, but they often struggle to visualize the scale, scope, and complexity of real-world process plants. Process operators, on the other hand, have a good awareness of operations but not a thorough theoretical understanding of the processes' logic. The vast majority of studies overlook the option of leveraging the same immersive virtual environment to fulfil academic and corporate instructional needs. Undergraduate chemical engineering students and process operators, for example, prefer visual explanations over traditional text-based learning [15]. Unlike traditional classrooms, virtual and digital learning environments can accommodate a diverse spectrum of student users. Younger students, lifelong learners, and students who are unable to physically access or afford traditional learning facilities can all benefit from virtual reality learning. The societal impact of the empowering component of virtual learning has yet to be determined. More study is needed to evaluate virtual reality's more significant educational value on the student user population.

Virtual reality use has several advantages in the classroom. Scenario studies to assess if the benefits outweigh the costs might be useful in this case. There are also social difficulties with virtual reality technology for education, such as the necessity for application deployment and additional skills, which may limit its adoption and use. Investigating the virtual lab safety course is the goal of this research in the chemical engineering design course to see if it can be utilized in undergraduate chemical engineering instruction. It will be compared to real-world laboratory experiments, and it will be documented using standard laboratory notebook formats. A technical test on the lab safety course will be administered before and after the virtual activities to assess the degree of learning. In addition, a mechanism for evaluating virtual reality's educational impact on undergraduate students will be established.

V. CASE STUDY: USING LABSTER VIRTUAL LABORATORIES AT ELMERGIB UNIVWERISTY' CHEMICAL ENGINEERING DESIGN COURSE

Virtual lab simulations enable students to do laboratory experiments while also examining complicated concepts and theories without needing to commute to a physical science lab. Labster VR lab simulations, as shown in figure 3, which demonstrate science at the molecular level, are available to students for free. They can then use their scientific knowledge and sophisticated laboratory equipment, as in figure 4, to real-life challenges involving chemical processes. 3D learning that is gamified in a virtual environment may be a laboratory or imaginary plains, with a captivating narrative and a scoring

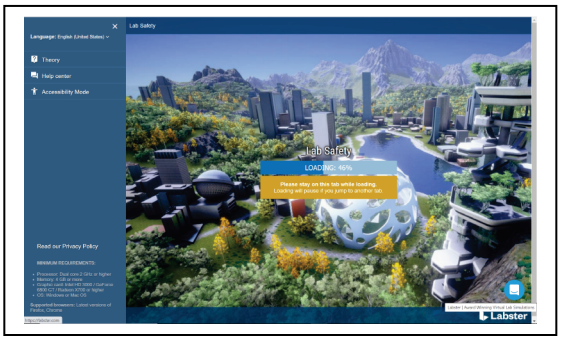


Figure 3: Labster VR lab simulations

Figure 4: VR sophisticated laboratory equipment

system. Labster provides virtual laboratory simulations in chemical engineering, biochemistry, biotechnology, chemistry, physics, and other Scientific areas; Labster can help students succeed and help students improve their learning outcomes. Labster users usually observe an improvement in average student grades. Improve student participation, and they choose to spend more than a few hours of their free time looking at a screen every day. Anyone may meet them in their natural environment and creatively teach them science with these realistic, 3D scientific simulations. The proposed VR tool uses gamification and storytelling strategies to keep students interested in the course material. The VR helps instructors to make students familiar with the lab, which may be frustrating and dangerous if students don't know how to interact with the science lab. The understudy VR lab can provide learners with a realistic virtual lab simulation before they do any genuine experiments in the real lab. That means students are free to make as many mistakes as they want and repeat their experiments as many times as they want with no negative implications in the real world. Personalize education for individuals with different requirements; some students may sail

through the lab modules while others struggle. The VR lab can help differentiate learning so that each student can work at their own pace. According to our research, students with lower levels of knowledge and ability tend to catch up with their peers while using the platform and practice as much as they can at their computer, as shown in figures 5 and 6. Instructors can save time by keeping track of the students' progress. Everything that

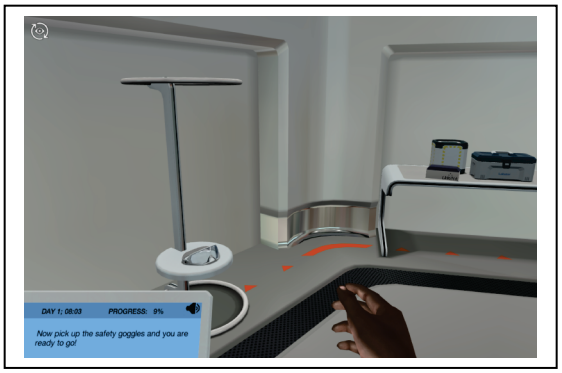


Figure 5: VR lab platform and practice instructed online

Figure 6: VR lab interactive with students simulating real lab

student does in this VR lab, including the number of quiz attempts, quiz question responses, time spent playing, and final grades earned, as in figures 7 and 8, is available to the course instructors and student group LMS as in Google Classroom.

Figure 7: VR lab interactive, guiding, and assessing students

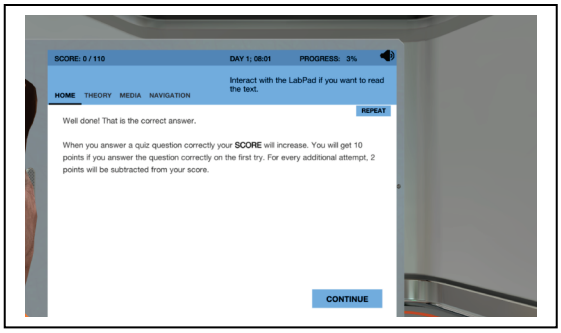


Figure 8: VR lab progress assessment with multiple questions

As an outcome, an instructor can be able to keep a continual check on the learners' progress and make adjustments based on behavioural data. Students should be taught to think in a scientific manner. VR lab is more than simply a laboratory simulator. It is the go-to teaching assistant, giving the participants the context they need to fully grasp the concepts instructors are teaching and make connections with their classmates. Instructors will be able to keep track of their student's grades and progress using the combination of the VR lab and the learning management system that can be connected with their dashboard. It allows for a comprehensive examination and analysis of students' efforts. At any point during the process, the instructors may see which simulations their students have started, their score, the number of attempts, and more.

Students' grades may be viewed in two ways: on the Grades page of Google Classroom or the VR lab dashboard. The dashboard will appear as an assignment in the Google Classroom course, as in figure 9, if the instructor adds a new simulation to the course work. When the course students play the simulations, the grades will be appropriately updated. The instructor can keep track of the class students' progress on the VR lab dashboard and monitor students' outcomes, as shown in figure 10. The instructors' dashboard allows them to monitor their students' grades and progress. It allows for a thorough review and analysis of students' efforts at any level of completion. Instructors can see which simulations they've started, their score, how many quizzes they've attempted, and when they finished them, among other things.

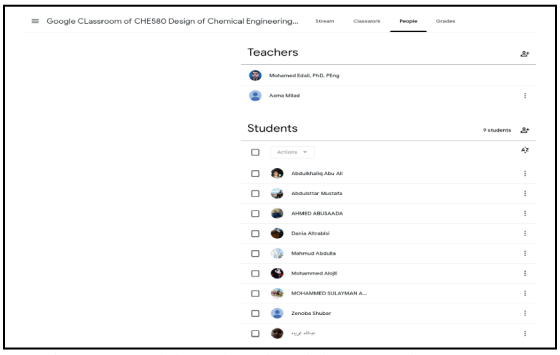


Figure 9: VR lab Assigned Training as Assignment #5 to Design Engineering Students at Elmergib University Google Classroom of CHE580 Course Portal.

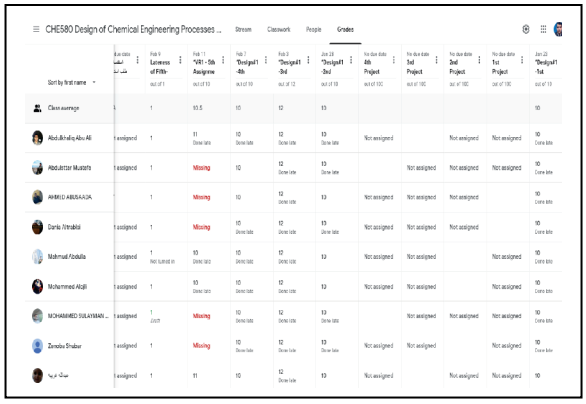


Figure 10: Google Classroom Grading Portal with the VR lab Assigned Training Assignments.

This study from Elmergib university's chemical engineering department examines the usage of virtual reality (VR) on nine registered fifth-year students to the design engineering course CHE580. This integrative review paper is a preliminary phase toward reporting to the scientific department committee on alterations to the course curriculum as engineering education development, with a focus on applications in academia and the process field. The study will evaluate the advantages and disadvantages of utilizing virtual reality in chemical engineering education, with an emphasis on the primary areas of design technology, pedagogy, and socioeconomics.

VI. Virtual Reality Teaching Issues Associated with the Chemical Engineering Education System

Adding mathematical models to virtual reality tools for advanced educational applications is universally acknowledged, enabling students to comprehend the conceptual framework as well as its outcomes in a virtual reality environment. Virtual reality is still in its early stage in respect of deployment and implementation in chemical engineering education and curriculum. The hereunder are amongst the prominent obstacles. Despite the fact that the implementation of such virtual reality instructional practises necessarily implies instructors with scientific awareness in video gaming design and development, as well as expertise in chemical engineering, undergraduate process modelling methodologies, and the acquisition of additional skills and resources. Faculty members in chemical engineering must collaborate with scientists and experts to guarantee certain VR technology will be used effectively in the classroom. The lack of access to virtual reality sophisticated unique hardware systems to undergraduates may jeopardize the professors' objective experiential learning.

VII. The Prospects of Chemical Engineering Curriculum and Instruction: Virtual Reality

Virtual reality's prospects for the future in chemical engineering and education are promising. Anticipate seeing a significant rise in the use of this technology in the near future. The developments point to VR's future progression. Due to the limits imposed by Corona, the development's methodology had to be modified. Instead, the educator utilized a VR programme and broadcast the VR visualization over Zoom or any other used teaching platform. The other instructors and learners may see the 3D displayed science using basic VR goggles and their smart cellphones, and the instructor orally communicated what was presented on the screen. Almost all students evaluated on the usage of a few VR trials were favourable when it came to recent technological releases, claiming to welcome the opportunity to explore new digital teaching tools. Even though, prior to this course at Elmergib University, most students had never explored virtual reality for instructional reasons. As a result, students had high expectations when it came time to participate in the VR trials. Students said that the VR technology was quite beneficial after the course projected lab training. Students' enthusiasm for learning more about the benefits and drawbacks of utilizing virtual reality to visualize chemical engineering processes stemmed from their notion that virtual reality technology may be helpful in their future jobs. Due to the possible VR and Zoom combination of the teaching pedagogy, instructors were able to teach without, or with a very little chance of, Covid spreading. Virtual reality is clearly a practical visualization approach, providing students with engaging, pleasant, and stressful environments in which they may engage and be creative. In university studies, the relevance of creativity in the chemical engineering curriculum has developed, as has the use of virtual reality.

VIII. Conclusion

This research paper has looked at the current level of virtual reality technologies in chemical engineering education. Interest is there in academia and industry in using virtual reality in chemical engineering. Various projects have attempted to construct virtual process plants that may meet both industrial and academic instructional demands. In the core fields of technology, pedagogy, and virtual reality-based education, the report recognized prospects and drawbacks. The study emphasized the need to use mathematical models to supplement virtual reality interfaces and addressed the modelling methodologies used in such systems. The report also addressed the critical pedagogical advantages and disadvantages of such schooling. In order to evaluate virtual-reality-based learning, the article emphasized the necessity for unique educational effect evaluation approaches. A case study was also provided to analyze the consequences and obstacles of the gradual adoption of virtual reality tools in chemical engineering education. It may be argued that during the COVID pandemic, virtual technologies have become more promising. In this light, virtual reality might be regarded as a powerful tool for remote learning in chemical engineering education. Such experience, it is acknowledged, cannot replace physical industrial field trips. Whereas a genuine experience is not feasible, however, such virtual technologies offer high-quality substitutes. The ultimate objective for conducting this research was mainly to explain a set of VR-based teaching modules used at Elmergib University, with an emphasis on the pedagogical concerns that were addressed throughout the VR models' use. In terms of content, there are three essential levels of instruction when using the VR case study methodology.

The operator must first learn about the many chemicals used in the process, the risks associated with them and how to avoid them, and the appropriate Personal Protective Equipment. Before beginning the task, this phase focuses on learning about the procedure's risks and requirements for safety. The learner is then given the opportunity to learn and practise the response

technique in accordance with the Standard Operating Procedure. The student controls the tools both manually equipment identification and via the control screen next to the tools presented. In a unique mode, emergency events are added to the simulation, and the student must detect them and execute the relevant Emergency Standard Operating Procedure to resolve the issue before it becomes a catastrophic accident. The authors also looked at the potential of virtual reality as a widespread data distribution medium, with the goal of using it in scientific visualization and practical engineering applications in the years ahead. The authors would like to start expanding VR as a beneficial tool in Elmergib Chemical engineering courses, starting with relatively high scientific visualization and the implementation of virtual accident scenarios to reduce laboratory mishaps.

Furthermore, while the current study demonstrated that the virtual intelligent online operator could provide authentic updates for students' learning, it is uncertain whether this type of VR training, which includes monitoring aspects, is as effective as traditional classroom settings. Students can acquire hands-on lab training virtually before they graduate from universities in Libya that don't have enough funds to operate engineering labs with the necessary instruments and chemicals. Future research should look into how effectively the VR feedback aspects of other engineering and science courses are built pedagogically and technologically. In addition, further research should seek to train students to collaborate well in groups of students in virtual reality environments. Finally, research into the short- and medium-term effects of incorporating virtual reality into other engineering courses for fundamental science learning is critical.

ACKNOWLEDGMENT

Professor Elkamel of Waterloo University granted academic support to the authors to use any software simulator needed to build instructional materials based on the simulation process. The efforts of the authors' creativity group in gathering all teaching tool specifications, as well as their time spent producing the majority of the work for the teaching website referenced in this study, particularly Mr. Sahem, are greatly appreciated. Elmergib University's chemical engineering department's associate professor Edali et al. 2021/2022 designed the teaching strategies and evaluated the utilized VR lab, as well as their teaching methodology and teaching philosophy videos in demand lectures and presentations and posted them to the course virtual google classroom. Students were then encouraged to use the teaching materials to practice on VR websites online on their laptops at their own pace. In addition, as shown in [16], authors launched a website for the *ILCICT-2022* and other publications into other conferences and the tools used as Apps and teaching techniques and know-How to operate and utilize, which any professor or engineering student can access.

REFERENCES

[1] Patle, Dipesh S., Manca, Davide, Nazir, Salman, Sharma, Swapnil, 2019. Opera-tor training simulators in virtual reality environment for process operators: a review. Virtual Real. 23, 293–311, http://dx.doi.org/10.1007/s10055-018-0354-3.

[2] Manca, Davide, Nazir, Salman, Komulainenc, Tiina, Øvergård, Kjell Ivar, 2016. How extreme environments can impact the training of industrial operators. Chem. Eng. Trans. 53, 193–198, http://dx.doi.org/10.3303/CET1653033.

[3] Fracaro, Sofia Garcia, Chan, Philippe, Gallagher, Timothy, Tehreem, Yusra, Toyoda, Ryo, Bernaerts, Kristel, Glassey, Jarka, et al., 2021. Towards design guidelines for virtual reality training for the chemical industry. Educ. Chem. Eng., 12–23.

[4] Ismail, Shahrul Mizan, Hashim, Harwati, 2020. Virtual Reality-Based Education(VRBE): understanding students' readiness and expectancies. Int. J. Innov. Technol. Explor. Eng., 172–176.

[5] Pirola, Carlo, Peretti, Cristina, Galli, Federico, 2020. Immersive virtual crude distillation unit learning experience: The EYE4EDU project. Comput. Chem. Eng.

[6] Wright, Davidson, 2020. How to tell the difference between a model and a digital twin. Adv. Model. Simul. Eng. Sci. 7, 13, http://dx.doi.org/10.1186/s40323-020-00147-4.153.

[7] Nielsen, Rasmus Fjordbak, Nazemzadeh, Nima, Sillesen, Laura Wind, Andersson,Martin Peter, Gernaey, Krist V., Mansouri, Seyed Soheil, 2020. Hybrid machine learning assisted modelling framework for particle processes. Comput. Chem. Eng.

[8] Lanzo, Valentine, Sohel, Yapp, 2020. A review of the uses of virtual reality in engineering education. Comput. Appl. Eng. Educ.

[9] Quintero, Jairo, Baldiris, Silvia, Rubira, Rainer, Cerón, Jhoni, Velez, Gloria, 2019. Aug-mented reality in educational inclusion. A systematic review on the last decade.Front. Psychol. 10, 1835, http://dx.doi.org/10.3389/fpsyg.2019.01835.

[10] Schofield, Damian., 2012. Mass effect: a chemical engineering application of virtual reality simulator technology. MERLOT J. Online Learn. Teach., 8 https://jolt.merlot.org/vol8no1/schofield 0312. HTML.

[11] Passos, Claudio, Nazir, Salman, Mol, Antonio C.A., Carvalho, PauloV.R., 2016. Collaborative virtual environment for training teams in emergency situations. Chem. Eng. Trans. 53, 217–222, http://dx.doi.org/10.3303/CET1653037.

[12] Kolomaznika, Miela, Sullivana, Miriam, Vyvyana, Kate, 2017. Can virtual reality engage students with teamwork? Int. J. Innov. Sci. Math. Educ. 25, 32–44. https://openjournals.library.sydney.edu.au/index.php/CAL/article/view/12172.

[13] Lin, Hsin, Wang, Hua, 2014. Avatar creation in virtual worlds: Behaviors and motivations. Comput. Human Behav.

[14] Ouyang, Shu-Guang, Wang, Gang, Yao, Jun-Yan, Zhu, Guang-Heng-Wei, Liu, Zhao-Yue, Feng, Chi, 2018. A Unity3D-based interactive three-dimensional virtual practice platform for chemical engineering. Comput. Appl. Eng. Educ. 26, 91– 100, http://dx.doi.org/10.1002/cae.21863.

[15] Norton, Christine, Cameron, Ian, Crosthwaite, Caroline, Balliu, Nicoleta, Tade, Moses, Shallcross, David, Hoadley, Andrew, Barton, Geoff, Kavanagh, John, 2008. Development and deployment of an immersive learning environment for enhancing process systems engineering concepts. Educ. Chem. Eng. 3, e75–e83, http://dx.doi.org/10.1016/j.ece.2008.04.00.

[16] Chemical Engineering Research Group website at Elmergib University for the ILCICT 2022, International Libyan Conference for Information and Communications Technologies, Mar 6th-9th 2022, Tripoli University, Libya.

https://sites.google.com/view/edali-et-al-engineering-educ/list-of-research-scientific-conferences-and-journals/ilcict-2022-mar-6-9-international-libyan-conference-for-info-and-com?authuser=0

&

https://sites.google.com/view/edali-et-al-engineering-educ/list-of-research-scientific-conferences-and-journals/jan-2022-international-conference-on-key-enabling-tech?authuser=0

Digital Transformation with Blockchain Technology: Applications and Research Directions

Asma Elmangoush
Collage of Industrial Technology
Electronic Engineering Department
Misurata, Libya
asma_elmangoush@cit.edu.ly

Fawzia Abujalala
Faculty of Information Technology
Misurata University
Misurata, Libya
f.abujalala@it.misuratau.edu.ly

Wafa A. Elshibani
Faculty of Information Technology
Misurata University
Misurata, Libya
w.elshibani@it.misuratau.edu.ly

Majdi Ashibani
Collage of Industrial Technology
Electronic Engineering Department
Misurata, Libya
majdi.ashibani @cit.edu.ly

Hana Eliwa
Faculty of Information Technology
Misurata University
Misurata, Libya
hanan_eliwa@it.misuratau.edu.ly

***Abstract*— The rise of digitalization has transformed the economy and business in various countries under the shade of technology. Libya, as many other countries, witnessed an incredible rise in technology that poses the way to a full digital transformation aiming to uplift the standards of living. In this paper, we discuss the role of Blockchain technology in digital transformation and present the researchers work different applications aiming to support the digital transformation in Libya. We argue that Blockchain technology could play a vital role in digital transformation within a Smart City vision, as well as e-government.**

Keywords: Digtal Transformation; Blockchain; Smart City, E-government.

I. INTRODUCTION

During the last few decades, there was an exponential increase in the usage of various smart applications such as smart healthcare, smart homes, energy management and smart grids etc. Besides the usage of the global Internet network for data transferring had enriched the provided features through the interoperability between different sectors. This paved the way for the digital transformation agenda all over the world.

However, for all the aforementioned smart applications, security and privacy are major concerns. Researcher and developers have put great efforts to propose new technologies and techniques to enhance the security levels in smart services. However, the existing solutions are either based upon the centralized architecture (having single point of failure) or having high computation and communication costs [1]. We argue here that Blockchain technology can be a good solution to overcome the limitations in existing smart systems.

The Blockchain is a novel disruptive technology based on cryptography. It has been known for the work of Nakamoto in 2008 that showed how this technology can become the core component to support transactions of the digital currency "bitcoin" [2] and address the double spending challenge for digital currencies. Additionally, Blockchain is a decentralized data structure and management technology developed to serve as a public ledger. With the introduction of Blockchain, many fields such as finance, accounting, and real estate will receive a positive impact using the benefits of this technology.

A Blockchain, is basically a growing list of records, called blocks, which are linked using cryptography. Each block contains a cryptographic hash of the previous block, a timestamp, public key of the next owner, and transaction data. By design, a blockchain is resistant to modification of the data. It is "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way". For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively proof-of-work adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Ownership verification is done by checking the signatures.

This paper provides a comprehensive view of the factors affecting digital transformation process and describes some applications to facilities digital transformation of main sector in Libya. The selected applications have been implemented as proof of concept by the team.

The rest of this paper is organized as follows: Section II discusses the digital transformation process and enablers technologies. Section III introduces the Blockchain technology. Section IV overviews three applications our team developed based on Blockchain technology from digital transformation. Section V discusses the future trends in Blockchain applications. Finally, Section VI concludes the paper.

II. THE PROCESS OF DIGITAL TRANSFORMATION

Initially, it is necessary to understand the difference between two basic terms associated with digital transformation: digitization and digitalization. These terms are often used interchangeably with digital transformation, but they refer to very different things. Digitization is the process of converting information from a physical format to digital format. It means converting something non-digital into a digital representation to be used by computer systems and automating processes or workflows. Digitization enables creating business value, which needs data. It helps to lay the foundation for business use cases that leverage the data management. For example, scanning a paper document and saving it on a hard drive as a digital document, like PDF format. While that, digitalization is the use of digital technologies and digitized data to impact how work gets done, transform how customers and companies engage and interact, and create new (digital) revenue streams.

According to McKinsey, digital transformation is rearrangement of technology, business models and processes as to ensure new values for customers and

employees in constantly changing and developing digital economy [3]. The COVID-19 crisis has brought about years of change in the way companies in all sectors and regions do business. New strategies and practices are required to stay competitive in the new economic environment. The following technologies could be considers as driving factors expediting Digital Transformation [4]:

1) Industry 4.0: with the fourth industrial revolution, it is foreseen that manufacturing process is digitized, machines are directly connected to each other and personalized manufacturing is possible.

2) Artificial Intelligence (AI): AI algorithms enables the digital system to learn, understand and judges itself. Thus, manufacturing could occur by robots having AI without human intervention and they are in continual learning and improvement.

3) Internet of Things (IoT): IoT is growing exponentially converting the physical world into a massive information system. It consists of interconnected devices from simple sensors to smartphones and wearable devices. IoT is an infrastructure open to new innovative information technologies.

4) Cloud Computing: a general term of internet based information services providing computation secrives in multiple levels.

5) Big Data: an evolving term used to describe any large amount of structured, semi-structured or unstructured data that has a potential to be mined for information. Big data help designers to derive decisive customer needs from the existing data to improve and develop designs.

6) Robotics: Robots contiutes to provide more exceptional ability and intelligence to automate tasks. Moreover, industries advanced high flexibility and shorter delivery time for products to the market.

7) 3D Printers: 3D printing belongs to a class of techniques known as additive manufacturing. Today, 3D printing can create objects from a variety of materials, including plastic, metal, ceramics, glass, paper, and even living cells. It's proclaimed that 3D printing will bring "the next industrial revolution." [5].

In order to get benefits from these technical drivers, companies and government authorities should transform their organizational structures and business making cultures starting from manufacturing technologies to management percepts as to get a productive digital transformation. The benefits from re-analysis the structure and processes are significant.

III. BLOCKCHAIN TECHNOLOGY OVERVIEW

Principally, Blockchain is a distributed database where assets can be stored and exchanged through a decentralized peer-to-peer network of computers. What distinguishes Blockchain from other databases is that the Blockchain ledger is an append-only database. Information in the Blockchain ledger cannot be changed or altered, i.e. every entry is a permanent entry. Furthermore, every entry is combined with the digital signatures of issuers. Thus, Blockchain technology provides good security for the user's data and also the transaction data.

The Blockchain ledger is a growing list of records, called blocks, which are linked using cryptography algorithms. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (Tx). Any attempt to change the data inside a block necessitates rehashing, not only the block relevant to the transaction but all subsequent blocks. This is possible theoretically, but it's impractical since the blocks grow continuously as other nodes add blocks to the Blockchain [2]. The data which is stored inside a block depends on the type of Blockchain application. For example, a Bitcoin block contains information about the sender, the receiver, and the number of Bitcoins to be transferred. The first block in the chain is called the Genesis block.

Each node in the Blockchain network is assigned two keys (Public key and Private key). The assets are distributed through the network, but only the owner who has the private key can make transactions on this asset. The other computers node in the network act as validators for the transaction (miners) [6]. Each transaction is verified for validity by the nodes in the Blockchain network, before recording it as a new Block into the Blockchain ledger. The nodes reach an agreement on which transactions must be kept in the Blockchain to guarantee that there will be no corrupt branches and divergences

There are many different blockchain network types. They can be distinguished depending on who can read, execute and validate transactions. Mainly, Blockchain networks are categorized into two types: public blockchain, and private blockchain. In public blockchain transactions are stored in different nodes in the distributed network, so it is nearly impossible to tamper with the public blockchain. In addition, propagate transactions and blocks as there are a large number of nodes on the public blockchain network. Taking network safety into consideration, restrictions on the public blockchain would be much strict. As a result, transaction throughput is limited and the latency is high. Moreover, everyone in the world could join the consensus process of the public blockchain [7].

As for the private chain, it is fully controlled by one organization that could determine the final consensus, while the read permission depends on a private blockchain. With fewer validators, private blockchain could be more efficient. The private blockchain is fully centralized as it is controlled by a single group. Different from public blockchain, a priviate blockchain is permissioned where one node needs to be certificated to join the consensus process in consortium or private blockchain [8]. Depending on the Blockchain type, different consensus mechanisms exist. The most well-known is the Proof-of-work (PoW). PoW requires solving a complicated computational process, like finding hashes with specific patterns. Interested readers could refer to [9] for a good survey on consensus mechanisms.

The Blockchain technology cannot only process currency transactions but can also ensure that transactions comply with programmable rules in the form of "smart contracts". A smart contract enables two or more parties to perform a trusted transaction without the need for intermediaries. How transactions are verified and added on the blockchain guarantees that conflicts or inaccuracies are reconciled and

that in the end there is only one valid transaction (no double entries). Many security researchers consider Blockchain as a promising solution to achieve a trusted Smart system, due to its capabilities such as immutability, transparency, auditability, data encryption, and operational resilience [10] [11].

IV. Applications for Libay Digital Transformation

Digital platforms as technical infrastructures and their ecosystems of social actors continue to change entire industries. New business models and ideas have evolved recently, utilizing the Blockchain's transparent and decentralized platform. In this section, we describe three applications that aim to enable the digital transformation in Libya.

A. Traceable Cargo System

Distributing heat-sensitive products (Drugs or vaccines) from manufacturer to patient is a very critical concern in many developing countries due to the hot climate and poor infrastructure. Cold chains define complex distribution processes by analyzing data constantly from the manufacturer until verifying the product to the consumer. This chain involves measuring, controlling, and documenting real-time values with absolutely no room for errors. Accordingly, cold chains should be monitored carefully to control the temperature and prevent broking the chain. For this project, we propose a traceable cargo system based on the Internet-of-Things (IoT) and Blockchain technologies. The use of IoT technology contributes to monitoring and recording temperature and humidity measurements through the whole chain. While Blockchain technology provides a decentralized, secure, and immutable shared ledger, which grantee data integrity. More important, the Blockchain's digital ledger is able to define the responsibility of any violation in the cargo in the supply chain.

In our use case, where vaccine packages are transported through a long road in different climatic conditions to consumers, the packaged vaccine shall be equipped with temperature and humidity sensors. The temperature and humidity data shall be aggregated at regular time intervals along the cold chain. To avoid any manipulation with the data, the Blockchain network will be in charge. A Smart Contract will be defined to control the operation of the whole system with reflection to the agreements between counterparties.

According to Henri [12], the details of the contract should always be publicly visible to all participants. During the transportation process, the current counterparty (CP) holding the cargo specifies the next CP responsible. The handover information along with the temperature and humidity data gets written inside a block and added to the Blockchain. This enables the cold chain owner (OW); usually the vaccine's manufacturer; and any cold chain observer (OB); the buyer and transporting companies; to figure out which counterparty is liable at any time. The system workflow in Fig. 1 indicates all possible states of the Smart contract, as well as the transition functions for each state. It also indicates when telemetry data is collected, and how the Smart contract specifics are enforced, should there be any humidity or temperature issues during the transportation process.

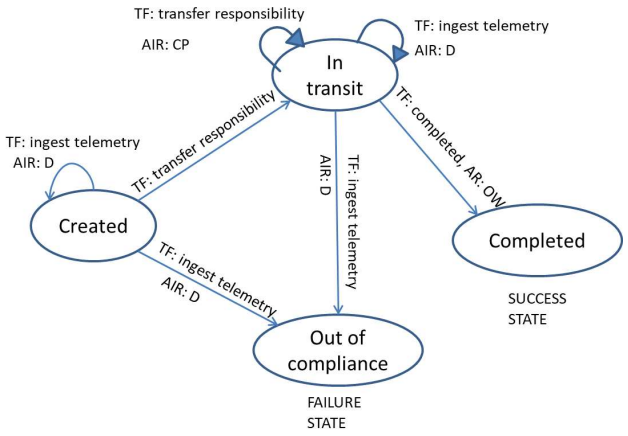


Fig. 1. Smart Contract Statues Diagram.

The Smart Contract is initiated in the *Created* state by the manufacturer (owner). From this point, the sensors are started to aggregate data to track the containers. Whenever the responsibility of the cargo is transferred, the state changed to *In transit* state. This state indicates which counterparty is responsible for the transferred goods (vaccines in our use-case) as well as ingesting telemetry. If the cargo reached its destination and contract rules have been verified, this will change the state to completed indicating that the smart contract ended successfully. However, if any of the contract rules have been broken during *Created* or *In transit* states, then the state will be changed to *Out of compliance* state indicating that the smart contract ended because of failure.

The system architecture adopted in this research is shown in Fig. 2. Each shipment consists of one container that contains a number of boxes. So, a smart sensor will be assigned to each box that communicates with a gateway in the container through Wi-Fi or Bluetooth. The gateway communicates with the IoT Hup to record telemetry data and device ID in a database. If the readings are out of the desired range, then it will be sent to the Blockchain network and the smart contract will be updated. An intelligent IoT gateway has been chosen to be implemented on a Raspberry-pi board to ensure M2M communication, as well as processing the data locally because it is not connected to the internet all the time. However, system implementation in real life has some

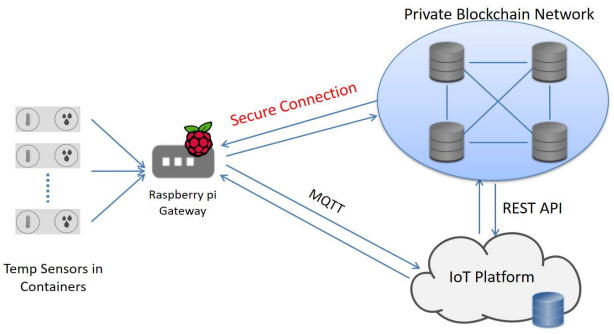


Fig. 2. Tracable Cargo System Architecture

challenges, as the IoT gateway could not communicate directly with the private Blockchain because of using the NAT server. This can be solved with the usage of IPv6 where all machines can have public IPv6.

B. Central Bank Digital Currency (CBDC)

The Central Bank Digital Currencies (CBDCs), or national digital currencies, are virtual currencies that are issued and controlled by the central bank (banking regulator). Hence, they are fully regulated by the state. CBDCs are not decentralized, like most cryptocurrencies; instead, they represent fiat money, only in digital form. Many banks around the world have started to investigate the new technology and test how to include it in their banking systems. Many considerations should be taken while proposing a fiat cryptocurrency, some of them related to technical aspects and others to monetary policy, financial stability, and legislation. Our team has developed, based on Blockchain technology, a Libyan digital coin (E-Dinar) that can be issued and controlled by the Central Bank of Libya (CBL).

CBL has been providing the public with cash banknotes and coins to use for payments. Cash has been considered as a preferred means of payment and has the confidence of Libyan citizens and facilitates trade in goods and services. Today's electronic and mobile payment markets are mainly in a situation in which payments that are intensively accessible to citizens /businesses are issued and controlled by private agents. If the state, via the central bank of Libya, does not have any digital payment services to offer as an alternative to the strongly concentrated private digital payment market, it may lead to a decline in competitiveness, cash availability, and a less stable payment system, as well as difficulties for certain individuals/business entities to make payments.

The current electronic/digital payment technologies in banking applications are still not ideal in many ways; firstly, the Libyan traditional electronic transaction systems require a manual intervention to work, it also based on trusted third parties. This will increase the underneath technology costs as it requires manual intervention work and periodic integrity checks [13]. Additionally, the current financial ecosystem, which is based on the behavior of its third parties, has proven short-sighted. Consequently, that has resulted in major crises in the global economy and led to a decrease in the level of trust in the ecosystem. [14] In addition to other economic challenges related to cryptocurrencies such as bitcoin, including money laundering, terrorism, price volatility, inflation, etc., Other problems are related to the Libyan case, such as liquidity problems and a lack of integration with e-commerce.

In this work, a pilot prototype was implemented and evaluated to develop the Libyan CBDC based on a private Blockchain network. Additionally, this research work has addressed the benefits and challenges that might arise during the running of such banking model. Figure 3 illustrates the proposed model scenario with four nodes (Commercial Banks and Central Bank) considered as validating nodes in this system, all nodes will share a common distributed ledger. As it is shown in Figure 3, the client sends a transaction proposal to the commercial banks, this transaction includes the identity of the client, the transaction payload, receiver public key and transaction id. The commercial banks will validate the transaction by checking the balance of the sender and its identity without updating the ledger. Then they will send the results in signed messages to the client that collects the

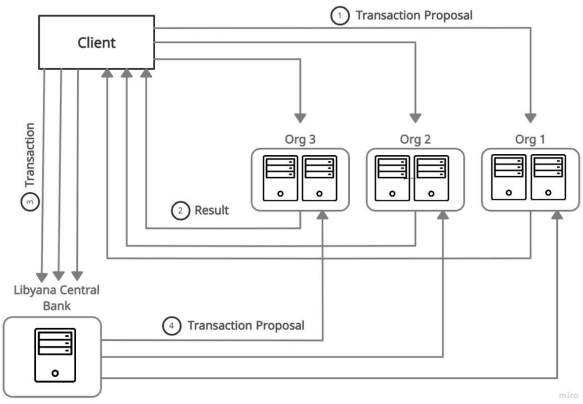


Fig. 3. E-dinar prototype model

required percentage of conciseness according to a predetermined policy. Once the client has collected enough correct responses, it will send the transaction to the Central Bank, which will collect and put them into blocks and then send them to commercial banks to be reflected and committed into the ledger.

The model of the Libyan dinar was designed based on one of the well-known frameworks known as CBCD two-tire model Figure 4, This model can also be modified in the line with the policies of the Central Bank or developing new policies to ensure the standards of safety and privacy on one hand and the national economic interests on the other hand.

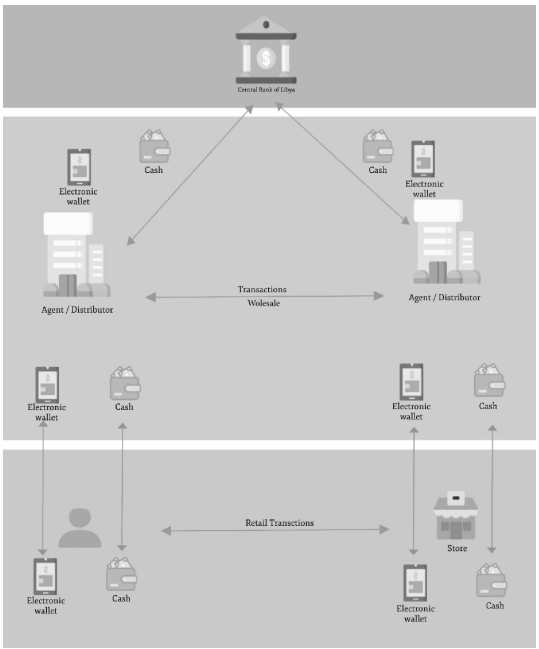


Fig. 4. Tow-tire CBCD model for E-dinar

The following are the working steps of the e-dinar prototype model, Figure 5:

1) The issuance department of the Central Bank issues the encrypted digital currency based on a decision of the Governor of the Central Bank of Libya according to policies of the country.

2) The Central Bank appoints some agents or distributors for electronic currency. These agents perform two basic roles; The first role is to create user accounts, this role will be dispensed with if there is a national electronic identity for citizens.The second role is to provide citizens with an electronic copy of the Libyan dinar in exchange for a paper copy of it, and vice versa. In this case, It's not necessary to have a bank account.

3) Upon a pledge request of commercial Banks, the central bank replaces the regular currency in the bank accounts with digital currency.

4) Transferring process of EDinar in trading will be done using mobile electronic wallets.

5) The settlement process will be done automatically within the same moment and the ledger will be updated too.

In addition, the team has developed a mobile application for the end users to login to their accounts and perform money payments. Figure 6 shows the app interface.

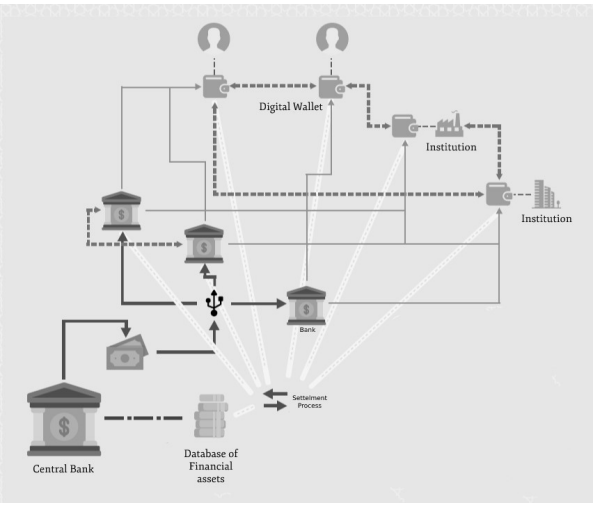


Fig. 6. The way how electronic currency model work

C. *Digitalization of Libyan Land Registry*

Land registration refers to a system whereby ownership and land-related rights are recorded by government authority. These records provide evidence of title, facilitate transactions, and prevent fraud. In Libya, the current land registry system suffers from many problems, being a purely paper-intensive system, the current traditional business processes are very lengthy and time-consuming. Moreover, other challenges of current business processes for government registry departments come from record issues that limit real estate stakeholders from obtaining, verifying or transferring legal proof of ownership.

Our team proposes a solution that could automate the registration process and reduce problems in current business processes at the Land Registry authority. The proposed solution is based on the Blockchain technology. It's expected that this solution will support the replacement of the current paper-intensive traditional land registration business environment, with a digitally secured paperless automated land registry for real estate transactions. In the analysis phase

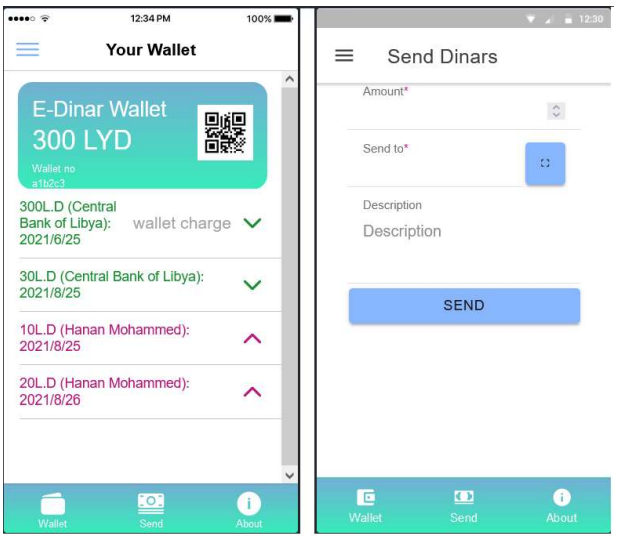


Fig. 5. E-dinar Application

of the research work, the business and legal processes (As-Is) for the Libyan Land Registry were documented and analyzed. These processes will be re-Engineered (To-Be) according to the Libyan Land Registry business and technology requirements. A key success factor in this step is the identification of:

- the entities (assets) and parties (participants) involved in these processes,
- their roles in each process,

and the relationships between them and their legal consequences according to the Libyan regulations.

As a proof of concept, we developed a prototype of the Libyan Land Registry system on top of Hyperledger Fabric platform [15]. A smart contract is implemented in JavaScript. The main components of the smart contract are state variables, functions, modifiers, and events. Typically, the main functions of the Land Registry are real estate registration and property transfer, so the asset to be traded in the blockchain network is the real estate. This real estate has an owner(s) who are participants in the blockchain model. Contract editors and Banks are also participants as a result of relationships between them and real estate. Moreover, Libyan Land Registry network consists of three nodes, each node has one peer. The nodes communicate with each other through a channel as they share a smart contract - chaincode. Every peer has copy of history log called a ledger depends on distributed ledger technology (DLT). Ladger contains all relevant data about transactions that occur within the network. In addition, certification authority (CA) is responsible for issuing PKI for peers.

The proposed network is exposed in Figure 7 to show how the client application (A) connects to the Blockchain network by requesting a transaction (query or update). The client requests a service through the application (A). Then the application (A) connects to a peer node) and invokes a smart contract to query or update the ledger. The peer invokes smart contract to generate a proposal response that contains a query result or a proposed ledger update. The client application (A) receives the response for the service request and queries. Then

the process will be considered complete. For updates, the client application builds a transaction from all of the responses, which sends them to the Libyan Land Registry for ordering. The Libyan Land Registry collects transactions from the network into blocks and distributes these to all peers. After that, the peers validate the transaction before applying it to the ledger. Peer generates an event, received by the client application (A), to signify completion.

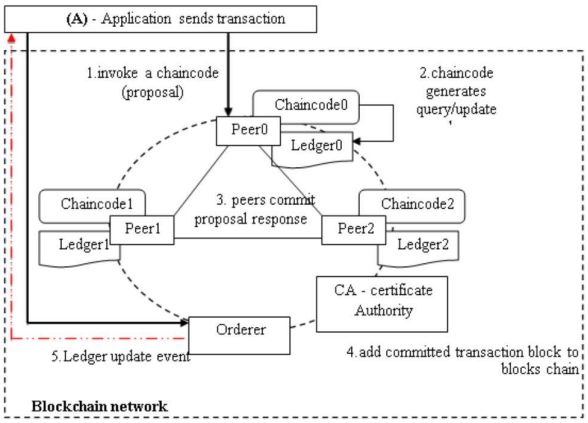


Fig. 7. Digital Land-Regitery Model Scenario

V. Blockchain Reeasrch Trends

Blockchain has shown its potential in industry and academia. We discuss possible future directions with respect to following areas:

1) Big data analytics: it's foreseen that blockchain will be used to store important data from varois applications. Blockchain could also ensure the data is original and secure. For example, if blockchain is used to store patients health information, the information could not be tampered and it is hard to stole those private information. When it comes to data analytics, transactions on blockchain could be used for big data analytics. For example, user trading patterns might be exteracted to predict potential trading behaviours and learn how to imporove serivces.

2) Consensus Algorithms: A review of Blockchain applications in practice illustrates trade-offs are indeed occurring between scalability, security, and environmental sustainability. In pracrice, the PoW algorithm has led to an arms race with ever-expanding energy consumption. Alternative consensus mechanisms, including proof-of-stake and delegated proof-of-stake, address these concerns. Moreover, ways to address the problem are being devised, and major blockchains such as Ethereum plan to migrate from proof-of-work to proof-of-stake in the future [16].

VI. Conclusion and Future Work

This paper presents our ongoing work in blockchain applications for digital transformation. We also integrated the blockchain ledger technologies with IoT infrastructure. This integration aims to improve security in IoT applications and provide transparency and traceability through the data flow between counterparties. In our work, we address three different applications and developed a proof of concept for each. A traceable cargo system was proposed to ensure the Libyan citizens with the validity of any product that is delivered through the cold chain. In order to implement this efficiently, all the participating machines have to be connected together with IPv6, which is not available yet in Libya.

Digital identities are common need to implement those projects and develop a successful digital transformation in Libya. This feature enables authenticating different people, organizations, and entities of various types. These digital identities can be issued by a governmental organization in a way like issuing driver's licenses, passports, or company registrations.

Acknowledgment

The work described in this paper was supported by the Al-madar Company, by hosting the developed services at Almadar-Cloud.

References

[1] U. Bodkhe *et al.*, "Blockchain for Industry 4.0: A comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020, doi: 10.1109/ACCESS.2020.2988579.

[2] R. Beck, "Beyond Bitcoin: The Rise of Blockchain World," *Computer (Long. Beach. Calif).*, vol. 51, no. 2, pp. 54–58, Feb. 2018, doi: 10.1109/MC.2018.1451660.

[3] McKinsey Glogal Institute, "How COVID-19 has pushed companies over the technology tipping point—and transformed business forever," 2020. [Online]. Available: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever.

[4] D. Ulas, "Digital Transformation Process and SMEs," *Procedia Comput. Sci.*, vol. 158, pp. 662–671, 2019, doi: 10.1016/j.procs.2019.09.101.

[5] J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs, "Disruptive technologies: Advances that will transform life, business, and the global economy," McKinsey Global Institute, San Francisco, 2013. [Online]. Available: http://www.mckinsey.com/insights/business_technology/disruptive_technologies.

[6] T. M. Fernández-Caramés and P. Fraga-Lamas, "A Review on the Use of Blockchain for the Internet of Things," *IEEE Access*, vol. 6, pp. 32979–33001, 2018, doi: 10.1109/ACCESS.2018.2842685.

[7] Z. Zheng, X. Chen, H.-N. Dai, and H. Wang, "Blockchain challenges and opportunities : a survey," *Int. J. Web Grid Serv.*, vol. 14, no. 4, pp. 352–375, 2018.

[8] K. Sultan, U. Ruhi, and R. Lakhani, "Conceptualizing blockchains: Characteristics & applications," *Proc. 11th IADIS Int. Conf. Inf. Syst. 2018, IS 2018*, pp. 49–57, 2018.

[9] W. Wang *et al.*, "A Survey on Consensus Mechanisms and Mining Strategy Management in Blockchain Networks," *IEEE Access*, vol. 18, no. 3, pp. 1–1, Apr. 2019, doi: 10.1109/ACCESS.2019.2896108.

[10] I. Makhdoom, M. Abolhasan, H. Abbas, and W. Ni, "Blockchain's adoption in IoT: The challenges, and a way forward," *J. Netw. Comput. Appl.*, vol. 125, no. September 2018, pp. 251–279, 2019, doi: 10.1016/j.jnca.2018.10.019.

[11] A. Panarello, N. Tapas, G. Merlino, F. Longo, and A. Puliafito, *Blockchain and iot integration: A systematic survey*, vol. 18, no. 8. 2018.

[12] C. GUTIERREZ and A. KHIZHNIAK, "IoT Meets Blockchain: Building a Supply Chain App on Microsoft Azure | Altoros." .

[13] H. Chen, X., Zhao, S., Wang, C., Zhang, S., & Cui, "GEEC: Scalable, Efficient, and Consistent Consensus for Blockchains," in *ArXiv e-prints (Aug. 2018) (2018)*.

[14] S. Nakamoto, "Bitcoin : A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[15] Hyperledger, "Hyperledger Fabric – Hyperledger." .

[16] M. Rossi, C. Mueller-Bloch, J. B. Thatcher, and R. Beck, "'Blockchain Research in Information Systems: Current Trends and an Inclusive Future Research Agenda,'" *J. Assoc. Inf. Syst.*, pp. 1388–1403, 2019, doi: 10.17705/1jais.00571.

Adopting Hybrid Requirements Modeling Tools For Web Applications Development

Mehemed Ben-Zahia, Ali Aburas, and Miloud Ghuwar
Computer Science Department
University of Tripoli
Tripoli, Libya
mohamedabdusalambenz1@gmail.com, aburasali@gmail.com, mghuwar@yahoo.com

***Abstract*—The web development process faces some obstacles, such as the tools used to model requirements. Developers face a dilemma of what type of modeling tools from different methodologies to support handling requirements in web developing projects. These tools include standard, object-oriented, Agile, and web tools. This paper illustrates different web requirements, such as context, presentation, navigation, and transaction requirements. In addition, the paper shows how modeling tools from diverse methodologies (Hybrid) can contribute to specific web requirements. The tools pool can be used for modeling web requirements. We believe these hybrid tools bring flexibility and ease to web application developers.**

***Index Terms*—Web Applications Requirements, Web Application Development, Modeling Tools**

I. Introduction

Requirement modeling is a technique used in software development to solve the requirement problem. The modeling is an important step in the software development [8]. Web requirement modeling is an approach for web development that uses tools to document web requirements in different forms such as natural language, use cases, and user stories. For example, to model a transaction requirement, we use a user story or a use case diagram. Developers need this approach to avoid misunderstanding of requirements taken from user side due to the lack of communication between the customer and the developer. The elicitation and specification of web requirements is a complex problem [3]. Many web development methodologies focused on user interface, but failed to address the overall development process of web applications [6]. The requirement modeling for web applications is different from generic applications in many aspects [11].

Web applications development requires a mix of web site development techniques together with properties of traditional software process models [20]. There are many tools to specify user requirements. These tools are taken from different methodologies, namely, Object Oriented, Structured and Agile ones. Agility can be applied to any software process [1]. When modeling web applications, we should take into account three levels, namely, content, hypertext and presentation [16]. Moreover, Object oriented tools do not provide appropriate concepts for the specification of hyperlink [16]. Use case diagrams are proven to be valuable in web development. The overall functionality of a web application is modeled as a set of use cases [16]. Use cases are the preferred modeling tool for web transaction requirements. In order to improve the analysis, user scenarios can be defined and used in the web process [2].

The rest of this paper is organized as follows: Section II gives a brief introduction into web requirement modeling, its then describes the challenges that affect the analysis and modelling of web applications; Section II-A to Section II-D lists the desirable properties of web application requirements modeling. Finally, a conclusion is given in Section III.

II. Adopting hybrid tools for web requirement modeling

Web requirement modeling is an important step in the web development process. Requirements are the backbone of any application and any deficiency in them would result in a negative impact on the outcome of web application project. It is the important activity of the web requirement engineering life cycle. Requirement modeling deals with the important elements of web requirements, namely, content, navigation, functional and presentation requirements. We concentrate in this paper on different types of web requirements and what tools that can be adopted to model them. These tools can be adopted from different methodologies, such as Object Oriented, Structured and Agile, to improve the outcome of the web application projects.

Many software process methodologies such as Agile and Object Oriented can be adopted and used to model some type of web requirements, especially the presentation and transaction requirements. Unfortunately, many web developers stressed on user interface, but failed to stress on other types of requirements. Traditional software process models have challenged to accommodate web specific aspects into their techniques and practices [1]. Souza and Falbo [19] proposed an Agile Approach for development of web applications which used the concept of Agile modeling nature of the web. When adopting traditional modeling tools to Web Applications, some aspects should be considered. These aspects include nature of the web project, such as size, application domain. For example, Agile methodology is suitable for small size projects. Requirements have many types explained below.

A. Content requirements

Content is king, a saying from the digital world. It means that content requirements are vital to the customer of the web application. There are many reasons of the importance of content. First, the content of the web application more important than the design. Second, the content of one web application can be more competitive to other web applications in the market. When developing a web application, content requirement plays an important role in the success of web products. Web content is mixed of visual, aural and textual contents.

Content is the elements on the pages of a web site. The elements of the content may include data, documents, video, audio and more. These different content types can be mixed together to form a single requirement. The data can be either structured or unstructured. Records of the structured form are of fixed size and can be modeled using structured tools like Entity Relationship Diagrams. ERD is the most common used of data models [13]. Class Diagrams [18] as an object oriented tool can be also used to model data which is an important element of content requirement of web application. The diagram is an illustration of the relationships among classes in the UML of the object oriented methodology.

B. Transaction Requirement

Transaction requirement deals with user needs or the functions expected from the proposed web application. It can be calculation of wage, authentication or data processing. Web transaction requirements can be expressed in the form of either graphical or textual manners.

Traditional methodologies can provide tools to specify web transaction requirements. For example, we can adopt Use Case Diagrams (UCD) and Scenarios from the Object Oriented methodology to model web transaction requirements. UCD plays an important role in modeling with UML.There are used for documenting system requirements [9]. The diagram is an Object Oriented analysis tool used to model functions of the system using actors and use cases. UCD is a graphical tool that consists of use cases as the operations performed by the system and actors as users or systems that perform these operations.

Use case scenario is another tool that can be used to model web functional requirements [17]. Each use case which used to define functional requirements of software system [4] has a description and this description is the use case scenario. For example, the use case cash a check of banking system has some steps in the form of procedure. This procedure is called use case scenario. A User Story is a tool from the Agile methodology that can be good choice to specify transaction requirements. User stories are popular methods for representing requirements [10]. A User Story is a narrative that illustrates how a user interacts with the system. It has the form: AS...I WANT...SO THAT...Structure English, from the structured methodology can be a good tool to specify process requirement.

All tools used in traditional software applications such as user story, Use Case Diagram , scenarios are applicable, but must include extensive collaboration of different types of stakeholders. A user satisfaction of the web application is very vital. Moreover, after modeling web requirements, they are subjected to extensive validation process.

C. Presentation Requirement

The web presentation requirement is concerned with the user interface, namely, the page layout that enables widely different users to communicate with the web application. This is a vital factor that contributes a lot to the success of any web application. The traditional methodologies contribute a little to the web methodology. Moreover, two difficulties arise here

1) Unlike classic software applications, users of a web application are not only the employees of the organization, but also their clients and maybe the public at large. Therefore, the layout pages of the application must be very appealing to most users. Moreover, users' involvement and participation in the design process makes the web application more acceptable. Feedbacks from different ranks of employees are of great help and different types of clients, and some other experts can provide positive contributions to an acceptable layout design. Using paper and dynamic prototypes from the traditional methodology, namely, the structured, would be of great help and should be considered.
2) Web layout design is not just a graphic user interface (GUI) as it in the case of classic software, where GUI is designed by programmers utilizing features of integrated development interface. However, in web applications, layout design requires multiple disciplines and specialists such as graphic designers and content writers. These specialists use software tools to create wireframes, and mockups to show them to the user. Wireframes et al. [7] provide a simple way for developers to create the basic design of an application. Sketching web applications with mockup tools is a common practice that improves the process of elicitation, and validation of requirements in web applications. Prototypes, also, from the classical methodologies can be of great help to model web requirements [14] [15].

User interface eases effective interaction between the visitor and the web application. The user requirements concerning presentation include ease of use, accessibility and visual aesthetic. The parts of the user interface include form based, touch, graphical and menu driven presentation. User interface requirement is the key to usability, which is one of the important attributes of web quality. The goal of effective user interface is to make the visitor of the web application more attracted.

D. Navigation Requirement

Navigation is vital to the success of any web application project. Because without navigation, the user would not be able to find his result or target, a blog or air fights for example. The structure of the navigation would allow us to visit any web page of the web application. Visitors of any web application

would stay if the navigation through the web pages are smooth and handy.

Web navigation requirement is one of the important factors to consider when designing a web application. However, web navigation is different from hyperlink. Hyperlinks are internal links throughout the content and are not considered part of the site navigation. Navigation does not have a large consideration in classic software systems, but highly emphasized requirement in web applications. Navigation is usually presented using a diagram that shows the structure of the web site [12].

Web application sitemap is a graphical tool that shows the structure and organization of the pages and content of the web application. This graphical tool provides linear and hierarchical page sequence but not random and direct page access within the navigation space of the web application. Sitemaps structure works best in web applications that requires direct access of information in a random order. It provides many navigation choices that users needs. Moreover, there is another tool from the management methodology that can be used to depict the structure of the web site.

Work Breakdown Structure Diagram (WBS) [5]is a management tool which helps to understand the required work needed to complete the web application project. The diagram is a graphical visualization in a hierarchical manner. It divides the project into components or tasks hierarchically. WBS can be used as a tool to model web navigation requirements.

III. Conclusion

Moving from classical software to modern web application products can not be done easily. It takes long and hard times to do this transition. Each product uses some tools for its requirements and design modeling of its methodology. To this end, we have proposed modeling tools adopted from different methodologies (namely, hybrid tools) to be used during the requirements analysis for web applications. Our objective and suggestion to adopt these tools to ease the web application process.

References

[1] OMAIMA NA AL-ALLAF. Hybrid web engineering process model for the development of large scale web applications. *Journal of Theoretical & Applied Information Technology*, 53(1), 2013.

[2] Davide Bolchini and Paolo Paolini. Capturing web application requirements through goal-oriented analysis. In *WER*, pages 16–28, 2002.

[3] M José Escalona and Nora Koch. Requirements engineering for web applications–a comparative study. *Journal of web Engineering*, pages 193–212, 2003.

[4] Javier J Gutiérrez, María J Escalona, Manuel Mejías, and Arturo H Zenteno. Using use case scenarios and operational vari-ables for generating test objectives. *Systems Testing and Validation*, page 23, 2007.

[5] Robert T Hans. Work breakdown structure: A tool for software project scope verification. *arXiv preprint arXiv:1308.2876*, 2013.

[6] Debra Howcroft and John Carroll. A proposed methodology for web development. pages 290–297, 01 2000.

[7] Aria YukFan Jim, Hyun Shim, Jue Wang, Lionel Richie Wijaya, Rongbin Xu, Hourieh Khalajzadeh, John Grundy, and Tanjila Kanij. Improving the modelling of human-centric aspects of software systems: A case study of modelling end user age in wirefame designs. In *ENASE*, pages 68–79, 2021.

[8] Yu Jun and Hu Zhi-yi. Requirements modeling of web-based scheduling information system. In *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*, volume 1, pages 105–108, 2008.

[9] Radoslaw Klimek and Piotr Szwed. Formal analysis of use case diagrams. *Computer Science*, 11:115–131, 2010.

[10] Garm Lucassen, Fabiano Dalpiaz, Jan Martijn EM van der Werf, and Sjaak Brinkkemper. The use and effectiveness of user stories in practice. In *International working conference on requirements engineering: Foundation for software quality*, pages 205–222. Springer, 2016.

[11] Neena Mishra and Garima Singh. Modelling of web based project using requirement engineering. *International Journal of technical Research and Applications*, 2016.

[12] Mark W Newman and James A Landay. Sitemaps, storyboards, and specifications: A sketch of web site design practice. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 263–274, 2000.

[13] MOHAMMED A Otair and AHMAD M Odat. Enhancing an end user development in database design using entity relationship diagram mapper. *Journal of Theoretical and Applied Information Technology*, 77(2):218–228, 2015.

[14] José Matías Rivero, Gustavo Rossi, Julián Grigera, Juan Burella, Esteban Robles Luna, and Silvia Gordillo. From mockups to user interface models: an extensible model driven approach. In *International Conference on Web Engineering*, pages 13–24. Springer, 2010.

[15] Thiago Rocha, Jean-Luc Hak, Marco Winckler, and Olivier Nicolas. A comparative study of milestones for featuring gui prototyping tools. *arXiv preprint arXiv:1906.01417*, 2019.

[16] Wieland Schwinger and Nora Koch. Modeling web applications. *Web Engineering*, pages 39–64, 2006.

[17] Michał Śmiałek, Jacek Bojarski, Wiktor Nowakowski, Albert Ambroziewicz, and Tomasz Straszak. Complementary use case scenario representations based on domain vocabularies. In *International Conference on Model Driven Engineering Languages and Systems*, pages 544–558. Springer, 2007.

[18] Alireza Souri, Mohammad ali Sharifloo, and Monire Norouzi. Formalizing class diagram in uml. In *2011 IEEE 2nd International Conference on Software Engineering and Service Science*, pages 524–527. IEEE, 2011.

[19] Vítor Estêvão Silva Souza and Ricardo de Almeida Falbo. An agile approach for web systems engineering. In *Proceedings of the 11th Brazilian Symposium on Multimedia and the web*, pages 1–3, 2005.

[20] R Vidgen. Wisdm: constructing a web information system development methodology. *Information Systems Journal*, 12(3):247–261, 2002.

Towards a Libyan Dialect Question Answering System

Ramadan Alsayed Alfared
Department of Computer Technology
University of Zawia- Faculty of Information Technology
Zawia, Libya
Ramadan.Alfared@zu.edu.ly

***Abstract*— The goal of this paper is to build a Libyan Dialect Question Answering System. QAS aims to satisfy users who are looking to answer a specific question in natural language. In additional, it is used to get most suitable and accurate answer to the questions asked by human in Natural Language. The result of this system contains a data set of 179 question statements with 135 of their answer that are written by Libyan Twitter's users as tweets. The dataset are extracting from Twitter with more 4000 statements. However, in the experiment of this study, we obtain 135 recognizing answers which are (75%, 60%, 71.6%, 81.9%, 50% and 75.8%) in (*who*, *when*, *where*, *what*, *how* and *why*) respectively. This result was stored in corpus, because no corpus of Libyan dialect question answer has emerged yet. Therefore, building such a corpus is very important to be used for the public purpose, not only for linguistic but also for the NLP research for Libyan dialect.**

Keywords—Answer Extraction, Information Extraction, Information Retrieval, Question Analysis, Question Answering.

I. INTRODUCTION

In the last years, Question Answering Systems (QAS) have become very popular and has been a topic of active research area in natural language processing (NLP) [1]. Arabic language or scripts come in many forms such as Modern Standard Arabic (MSA) [2], Classical Arabic (CA)[3] and Dialectal Arabic (DA)[4][5]. The Arabic dialects have become increasingly used in social media networks such as Twitter, Facebook, etc. Libyan dialect is one of many Arabic dialects which are spoken in North Africa. Nowadays, twitter social media network has hundreds of millions of active users all over the world, and about 500 million tweets are posted per day. Consequently, twitter is considered a rich and precious resource of data for researching. Therefore, the Question Answering (QA) has been one of the main focal points of research in the area of natural language processing such as Machine translation [6], Document classification [7] and Information extraction [8]. One of the most important and strongly needed *Question Answering* is a question analysis which is the first stage in any question-answer system and the accuracy of the results can significantly affect the subsequent stages of information retrieval and answer extraction. To get an efficient result, some semantic information may be available in questions that should be extracted for question analysis. Obviously, the most important step of question answering process starts with a question analysis phase that tries to determine what the question is looking for and how to effectively approach answering it. In our method, we use a set of pattern-template matching rules which can be sufficient to get a rather good result for question answering system in the accuracy level.

In addition, the question is expanded by adding synonyms of its terms "by using regular expressions" to improve the accuracy of the retrieval answers. After the question pre-processing and processing steps are done, the final stage is to find the answer from the retrieved tweets.

The request is usually in the form of a list of keywords. However, the system must go through the trouble of searching for the exact answer to their question inside each of the retrieved tweets. Nowadays, a new approach matches the needs by analyzing the question posted in the search field from a linguistic point of view, attempting to understand what the Twitters really mean and extracting the correct answer to the question from the retrieved answers.

The rest of the paper is organized as follows. Related work presents in section II. In section III we outline the essential step of collecting data corpus from the Social Media. Section IV explains the data preprocessing and corpus format. Section V contains an overview of the system architecture of QA system and detailed description of each system component.. Section VI presents the results of statistical analysis of corpus data. Final section concludes the paper and future work.

II. RELATED WORK

In the recent years many researchers have extensively studied and have tried to find clues to this problem in Arabic. In [10], they represent the answer extraction module encompass of three sub-modules: answer similarity checker, answers ranker and answer keywords stemmer. The answer similarity checker measured the similarity by counting the number of matching keywords between the question and retrieved answer. The answers ranker sorts answers according to their relevance. Then, return the top relevant answer. The answer keywords stemmer returns the roots of the keywords in the retrieved answer.

In [11], they have described an approach to the construction of a question answering system that provides short answers to questions expressed in the Arabic language. The system utilizes techniques from Information Retrieval and Natural Language Processing to process a collection of Arabic text documents. The overall success of the system is limited by the

number of available tools developed for the Arabic language. Work is undergoing to get retrieval integrated into the system and to extend the functionality of the NLP system by developing more sophisticated algorithms to produce a concise answer in a timely manner. Mostly, QA systems use different NLP techniques such as: super vector machine to classify questions, and lexicon based named entity recognizer to obtain the right answer [14]; others use surface patterns to extract important terms from questions, construct the relations of terms from sentences in the corpus, and then use these relations to filter appropriate answer candidates. The statistical model mostly requires the use association rules and is built by employing unsupervised learning of context patterns that indicates whether there is a presence of a question

III. SOCIAIL MEDIA DATASET EXTRACTION

For corpus collection, we use the Twitter search Application Programming Interface (API) called streaming API, which allows obtaining a stream of real-time tweets and sets of tweets from the past up to the last seven days by querying their content. In order to retrieve a collection of relevant tweets matching a specified query for the Libyan dialect, we create a set of search queries to increase the accuracy of obtaining tweets that are likely to be written in the Libyan dialect. For more details see [9]. The data is written by Libyan Twitter's users during the period from 1st Sep, 2021 to 30th Nov, 2021.

IV. DATA PRE-PROCESSING FORMAT

Unimportant text removal is an important step that should be considered during the preprocessing stages. In these stages, we developed a tool to corpus pre-processing, it was used for eliminating all unimportant text from tweets. For example, all of the following Twitter @-mentions, emoticons, #hash-tags and URLs were separately removed from each tweet by a tool in a preprocessing chain. After that, we had to remove some repeated sentences automatically. The remained tweets were organized into sentences. Then we decided to store sentences and all relevant data in a way that would be easier to handle in terms of data exchange and storage format. To do this, corpus data was stored in two data type format. Firstly, all data of corpus was stored in a relational database (MySQL) while the second type of data format is written in standard XML (Extensible Markup Language) format.

We properly specified the corpus structure in standard XML format to be readable for both human and machines. The current version of our XML corpus comprises one major element named <sentence> as well as it contains the attribute named *id*, the main element named <text> which has two child elements named <word> and <tag>.

V. SYSTEM ARCHITECTURE

The proposed method uses a set of pattern-template matching rules and statistical model with morphological analysis techniques "Part-Of-Speech Tagger" [12] for answers extraction. We also depend on a lexicon based named entity recognizer to obtain the right answer [13]. The workspace distiller which presented in Figure 1 takes an input file which contains the sentences with its POS tags (annotated sentence), and the final output is a file with questions and list of answers that is classified in the most appropriate answers.

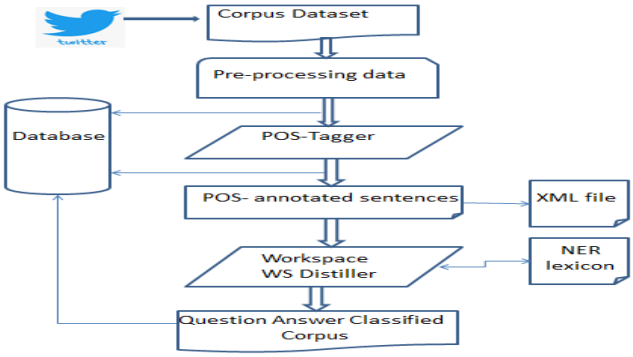


Fig. 1 :System Architecture

In most the languages, the question mark is typically placed at the end of a sentence, where it replaces the full stop (period). However, the question mark may also occur at the middle or end of a clause or phrase (see Figure 2), where there is no exact regal in Libyan dialect or like the other dialects. Our idea is to read the answer which is coming in a separate line after question mark and concatenate the answer lines till the next question mark found. In final we add the answer as a template. For each pattern obtained "*question terms*" from the previous phase, we check the presence of the pattern in the question sentence and try to classify them in the most appropriate answers.

Fig. 2: Different question types

Furthermore, the question mark can not be relied upon, as there are users whom write this mark at the end of the sentence and it does not express a question. On the contrary, we can find a sentence that does not contain a question mark and it is an interrogative sentence such as: sentences *id="2126"* in Figure 2. For example, if the question begin with or contains "وين", "where"; the question type will be location therefore, we need to look for all the possibility of *((DT)?(NN(s)?)* tags and also use the NER lexicon which has been built by the same author [11]. Table I represents the list of question keywords.

Table I: Question class and it definition

Question containing with	Question Class	Definition
من،مين، اشكون	Who, Whose	Person, Group
متى، امتى	When	Date, Time
ايذا، شن،شنو	What	Organization
وين، فين	Where	Location
قداش،كم	How Much, How Many	Number, Quantity, time
ليش، ع ش. ، ش	Why	Reason, cause

There are other types of questions that do not have a question mark, which is called an indirect question. Indirect questions are embedded within declarative statements: see Figure 2 in the sentence that has *id="2303"*. Many inflectional and derivation forms as well as it have many dialects sentences which require advanced techniques to solve such as training dataset and the statistical model.

VI. Statistical Results and Discussions

We built a program that applies all rules described previously using Php and MySql Database. The statistical results show some statistical information about the question answering system for Libyan Dialect. Figure 3 represents a question that contains the template "علاش" meaning *"why"* which will have two answers.

Fig. 3: Question with their answers

These two answers contain information "template "زواج" means *marriage*" that would be needed to determine the answers of the question of the sentence *id="3180"*. Figure 4 contains the word "شنو" meaning *"what"* in the sentence *"id=3114"* which will result in two answers in the sentences "*id=3127"* and *"id=3131"*.

Fig. 4: Samples of Libyan Dialect questions

Overall, the number of answers in the six question types shows considerable diversity. However, there was an upward trend in the number of both *Where* and *What* question, whereas there was a downward trend in the number of *who* and *when*. The following table summarizes the results obtained.

Table II: System evaluation of the six question types.

Question containing with	Question class	Total number	Correct answers	Precision
من،مين، اشكون	Who, Whose	4	3	75%
متى، امتى	When	10	6	60%
ايذا شن، شنو	What	60	43	71.6%
وين، فين	Where	72	59	81.9%
قداش،كم	How Much, How Many	4	2	50%
ليش،علاش، لاش	Why	29	22	75.8%
Total		179	135	75.41%

From the above table, some interesting observations can be made. For example, the type question *where* was the best results compared to the other question types, because we relied on the NER lexicon which improved the results.

VII. Conclusion and Future Work

In this research, we have presented a method to the construction of a question answering system that provides answers to questions expressed in the Libyan Dialect. The system utilizes techniques from *Information Retrieval* and *Natural Language Processing* to process a collection of a Libyan dialect as its primary source of knowledge for this domain. We used a set of pattern-template matching rules that achieve good results in spite of some errors that were

produced by the tagger. The results of our method would be enhanced if the tagger result was improved. In the future work, we will extend the result that is stored in corpus dataset, and we plan to use this dataset to train a new machine learning question answer systems to be the first application for Libyan dialect.

REFERENCES

[1] A Turing." Computing Machinery and Intelligence", Oxford University Press on behalf of the Mind Association, Mind 49, vol 59, 1950, pp. 433-460.

[2] I. Guellil, H. Saadane, F. Azouau, B. Gueni and D. Nouvel, "Arabic natural language processing: An overview", Journal of King Saud University-Computer and Information Sciences, Volume 33, Issue 5, June 2021, pp. 497-507.

[3] A. M. Ezzeldin, M. H. Kholief, and Y. El-Sonbaty, "Alqasim: Arabic language question answer selection in machines", in International Conference of the Cross-Language Evaluation Forum for European Languages. Springer, 2013, pp. 100–103.

[4] H. Mubarak and K. Darwish, "Using Twitter to collect a multi-dialectal corpus of arabic". In the Proc Of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), 2014, pp. 1–7.

[5] W. Zaghouani, A. Charfi, "Arap-Tweet: A Large Multi-Dialect Twitter Corpus for Gender, Age and Language Variety Identification". In Proceeding of the Eleventh International Conference on Language Resources and Evalution, 2018.

[6] F, Sebastiani, "Machine learning in automated text categorization", ACM Comput. Surv. 2002, pp. 1–47.

[7] L. Lulu, A. Elnagar, "Automatic Arabic Dialect Classification Using Deep Learning Models", Procedia Computer Science, Volume 14, 8102, pp 262-269.

[8] S. Abuleil, "Extracting Names From Arabic Text For Question Answering Systems", 7th International Conference of Computer- Assisted Information Retrieval Applications, University of Avignon, RIAO, France 2004.

[9] H.Alhami and R. Alfared, "Building a Twitter Social Media Network Corpus for Libyan Dialect", International Journal of Computer Electrical Engineering, Volume 10, Turky, 2017, pp. 558-564.

[10] H. Kurdi, S. Alkhaider and N. Alfaifi, “Development and evaluation of a web based question answering system for arabic language”, Computer Science & Information Technology (CS & IT), vol. 4, no. 2, 2014, pp. 187–202.

[11] G. k, A. Hammouri, R. Al-shalabi and M. Swalha, "A new Question Anszering Syste; for the Arabic Language", In American journal of Applied science, Apriel 2009.

[12] M. Oakes, J. Tait, “Word Sense Disambiguation in Information Retrieval Revisited”, in Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003, pp. 159-166.

[13] R. Alfared and H, Alhamim, "Libyan Dialect Named Entity Recognition On Rule-Based Approach", The Libyan International Conference on Electronic Engineering and Technology, LICEET 2018, March 3-7, Tripoli 2018, Libya.

[14] D. Zhang and W. Sun Lee, "Question classification using support vector machines" in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, July 2003, pp 26–32.

Efficient A-MPDU based on IEEE 802.11n WLANs

Ali Ahmad Milad
Computer Department
Faculty of Education,Elmergib University
AlKhums / Libya
alimilad@elmergib.edu.ly

Mustafa Almahdi Algaet
Computer Department
Faculty of Education, Elmergib University
Alkhums/ Libya
malgaet@elmergib.edu.ly

Saad Mohamed Lafi
Computer Department
Faculty of Education, Elmergib University
Alkhums/ Libya
saad.lafi@elmergib.edu.ly

Alhadi Ali Klaib
Department of Software Engeneering
Faculty of Information Technology, Elmergib University
Alkhums/Libya
alhadi.klaib@elmergib.edu.ly

***Abstract*—** In IEEE 802.11n, the Aggregation scheme of MAC Protocol Data Unit mainly connects multiple MPDUs to a single main PHY header. A-MPDU is failed to support error recovery; hence, if a single bit loss occurs during the transmission, then the entire frame will be transmitted leading to the overhead. Thus, in this study, a new scheme is proposed to tackle the problem called Fragmentation and Aggregation MPDU which aggregates packets and frames into a large frame. Each data frame is then divided into subframes, and each subframe is sent. In this scheme, if any errors occur during the transmission, only the corrupted subframes are retransmitted. The fragmentation method, in which the packets longer than a threshold are divided into fragments before combination, is also studied. The system is tested by NS-2 simulation. Simulation results show that the scheme significantly improves the performance of throughput and delay over normal A-MPDU. The results also reveal that the fragmentation scheme plays a key role in eliminating retransmission and erroneous overhead over noisy channels.

Keywords- IEEE 802.11n, MAC protocol, A-MPDU, Fragmentation Scheme

I. INTRODUCTION

WLANs have become the mainstream nowadays. WLAN success is due to its usefulness and efficiency in enhancing the rates of data transfer while keeping costs low. Considerable developments have recently been achieved in IEEE802.11 WLAN [1], widely accepted as a technology complying with IEEE 802.3 (high-speed Ethernet) for handheld and portable gadgets. IEEE 802.11, 802.11b, and 802.11a/g have data speeds of up to 2, 11, and 54 Mbps, respectively [2]. The IEEE 802.11 Working Group also introduces IEEE 802.11n, an updated version of IEEE 802.11 with improved throughput and speed. There is a difference between the performance of IEEE 802.11a, 11b, 11g, 11e, and that of 11n. Although all of them provide high data rates and a variety of physical layer (PHY) requirements, IEEE 802.11n provides a high throughput rather than high data rates, as well as improvements in PHY and medium access control (MAC). Moreover, the high efficiency in IEEE 802.11n enables it to cut down on the MAC layer overhead.

One of the features of frame aggregation is that, for transmission, it combines many data frames into a single large frame. After moving the data block from the sender to the receiver, a block of data frames known as block acknowledgment (BA) is sent sequentially from the destination. To identify sorts of frames taken by the receiver, the BA request (BAR) is conveyed from the sender. BA is then returned to the sender [3]. Physical data rates reached up to 600 Mbps were mentioned in most IEEE 802.11n proposals. However, since augmenting PHY rates with MAC layer overhead decreases efficiency significantly, these high PHY rates do not always lead to equal augmentation throughput especially at the upper layer; however, this phenomenon is a key function [4]. In order to ensure high-performance throughput for highly developed multimedia applications, increasing the MAC layer's performance is still inadequate [5]. Moreover, packet aggregation causes corrupted frames with large frame sizes, such as A-MPDU frames, resulting in considerable overhead when the channel is noisy. In the traditional transmission scheme, only one missing part can lead to the retransmission of the entire data frame [5].

The main goal of this paper is not only to reduce the overhead but also to speed up the efficiency of very high-speed WLAN IEEE 802.11n. Collisions and errors, which cause the DCF of the MAC layer to retransmit overheads, are the main causes of IEEE 802.11 throughput performance inefficiency in wireless networks. The current transmission using A-MPDU frames wastes majority of the channel time, avoids throughput scaling with PHY data rates, and decreases timing and header overheads. The increase in throughput is proportional to the increase in data rate. Transmission errors and collisions, however, continue to cause overhead retransmissions [6]. This research aims to reduce the retransmission overhead caused by transmission errors and collections.

II. BACKGROUND AND RELATED WORKS

DCF is a CSMA/CA mechanism used in the 802.11 MAC is DCF. Before implementing a transmission in a station, a clear channel assessment is performed in DCF IFS (DIFS). The station believes that the ownership of the medium has been taken over and a frame exchange is started in sequence when the medium is inactive. During station engagement, the station awaits the medium to go into inactivity mode, then defers to DIFS, followed by a random back-off duration. In the presence of inactivity, the station not only assumes medium ownership but also begins a frame exchange sequence for DIFS deferral and back-off length using the minimum distance and short IFS (SIFS) between

frames in sequence maintenance. Moreover, medium control is retained when stations obtain access to the medium [7]. Since a fixed time longer than SIFS must be postponed, another station would not be able to access the medium via this series. To prevent one station from having a monopoly on the medium, rules restrict the frame exchange sequence types permitted as well as their length. Several stations will wait for the medium to become inactive. During the medium engagement, the network is loaded by sending assembled packets. The random collection of back-off distinctive intervals by each station results in a low likelihood of collisions, when transmissions are initiated by more than one station [8], [9].

Because of the latent time and association of repetitive processes, data transmission over a wireless medium is unaffected. The process stimulates frame retransmission when the recipient's demodulation fails. To achieve the aforementioned method, it is necessity to have a sort of receiving station for the receiving data frames as well as sending a simultaneous plus positive ACK frame recognition. In the absence of an ACK frame at the data sending station, data can be retransmitted [10], [11].

The data transfer through the Data/ACK exchange is depicted in Figure 1. The data is transferred from STA1 to STA2 as shown in the figure. STA 1 accesses the medium, which then, terminates for DIFS, and then, goes after a random back-off period. If the medium remains in an inert state, data is retransmitted to STA 2. Moreover, STA 2 detects and demodulates the frame, which is followed by an ACK response. STA 1 then embraces the frame and re-accesses the channel in preparation for the next transmission. STA 1 will not receive ACK If SAT2 fails to demodulate the frames, and retransmission will be attempted again for the next transmission [12].

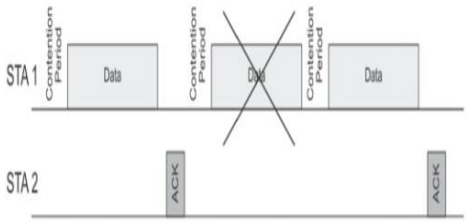


Fig. 1: Basic Data/ACK Frame Exchange Sequence

The IEEE 802.11 Working Group has adopted IEEE 802.11n, a new version with higher throughput level and acceleration. IEEE 802.11n, which includes PHY and MAC enhancements, has a higher throughput standard rather than a large data frequency [13]. Furthermore, IEEE 802.11n features frame aggregation, Block ACK BA, and reverse direction transmission, all of which help to reduce MAC layer overhead [14],[15],[16]. Frame aggregation is a technique for merging several frames into a single large individual one for transmission. The operation is performed with the help of two existing procedures: aggregate MAC services data unit (A-MSDU) and aggregate PDU (A-MPDU). Moreover, when it comes to information transfer features, MSDU can be differentiated from MPDU. MSDU refers to the import or export of data from the highest MAC layer, while MPDU depends for the lower section of MAC on the transmission to or from PHY [17]. Multiple MPDU subframes are linked to a single main PHY header by an A-MPDU aggregate. When it comes to function, A-MSDU differs from A-MPDU since A-MPDU begins after the MAC header enclosure procedure. Nevertheless, dealing with all of the MPDUs bound by A-MPDU includes the use of equivalent recipients. Furthermore, creating an A-MPDU necessitates some waiting/holding time, and the number of MPDUs required is solely determined by the number of packets already present in the transmission sequence, as shown in Figure 2. Furthermore, A-MPDU has the highest capacity and the PSDU has the highest capacity that is needed to be accepted, both of which are 65,535 bytes. However, the value can be reduced on the basis of existing STA capabilities in the HT elementary capability. Moreover, due to the capability of the unit ACK bitmap (128 bytes), in which 2 bytes are employed in order to map each frame required to acknowledge up to 16 [18], [19], the maximum acquired value of the subframes is 64.

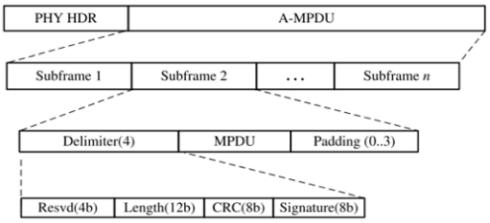


Fig. 2: Frame of A-MPDU

Before the implementation of aggregation in the 802.11n amendment, the block ACK protocol was proposed. A PHY transmission, like the original Data/ACK mechanism, consists of several MPDUs. During aggregate transmission, the mechanism proposed in the 802.11n amendment supports the improvements in the Normal ACK policy in the QoS frames. The BA is returned by the responder in response to the aggregate when the ACK Policy area of one or more QoS Data MPDUs in an aggregate is set to Normal ACK. The two existing response mechanisms that are modified to the Normal ACK policy can be observed in Figure 3 [20].

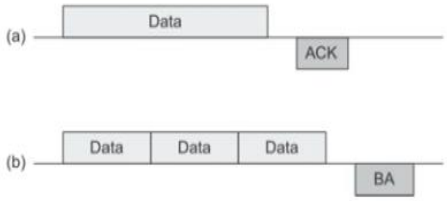


Fig. 3: Normal ACK policy and aggregate

The proposed BA protocol also enhances performance if the transmission of a block of data frames is permitted. Each data frame is recognized with a single BA frame instead of an ACK frame for each frame of the data. Compared with the standard acknowledgement mechanism, BA has a session-oriented mechanism. For each TID, a station needs to create a BA session along with a peer station so that the transfer of block data can take place. As shown in Figure 4 [21], the transmit address, receiver address, and TID all function together to define a specific BA session.

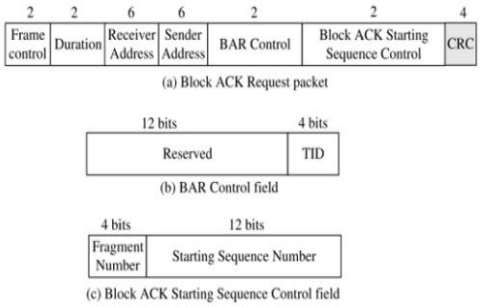


Fig. 4: Format of the BAR frame

III. Design And Methodology

The proposed scheme is primarily concerned with the A-MPDU frame. The scheme improves the current A-MPDU by making it more robust and efficient in a variety of ways. When robustness and efficiency are achieved by aggregation, fragmentation, and retransmission of the corrupted subframe, the integrity check is allowed. FA-MPDU operates by merging frames and packets into a superior frame, separating it into subframes, and then transmitting each subframe. In the event of a transmission error, only the corrupted subframes should be retransmitted, rather than the entire frame. The fragmentation process, in which packets longer than a threshold are separated into fragments before the combination, is also investigated. The FA-MPDU scheme was used in single hope topology in NS2. All stations involved in a transmission session are put in the same transmission range with enough transmission power to reach all other STAs using the TCP protocol. In the network, hidden terminals do not exist. Variety of channel measurements are used to examine the scheme performance, including the number of transmission stations, fragment capacity, and bit error rate. In Table 1, the parameters used in NS-2 have been explained. This simulation was created to examine the throughput and delay of the scheme in noisy channels.

TABLE I: NS-2 Parameters Used in FA-MPDU Scheme

Parameters	TCP
Number of stations	50, varied
Application rate	N/A
Data rate	648 Mb/s
Basic rate	54 Mb/s
Packet size	1024 B
DCF frame size	1024 B
A-MPDU, FA-MPDU frame size	65,536 B
FA-MPDU subframe size	varied

A-MPDU frame format

The FA-MPDU scheme operates by splitting a data frame into subframes, then every subframe is transmitted separately. In cases of errors during the transmission, the scheme just retransmits the corrupted subframe rather than the whole frame. According to the IEEE 802.11n standard, the maximum data frame size is 65,536 bytes. As a result, eight separate data frames were combined into a single wide frame used in the FA-MPDU scheme. Figure 5 shows the FA-MPDU frame.

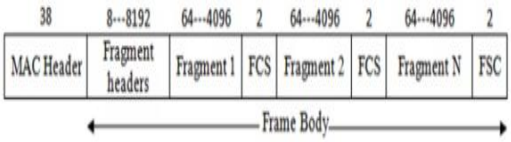


Fig. 5: FA-MPDU frame format

A BA bitmap field is included in the BA frame, as shown in Figure 6, to assist the transmitter in determining which subframes have been lost in a data block and determining successful and unsuccessful subframes. The bitmap should be 128 bytes in size to accommodate all of the data fragment figures. When each subframe is 128 bytes in size, the maximum number of subframes in the FA-MPDU scheme is 512. A 512-bit bitmap is used, which ensures that a wide block will hold up to 128 × 8 = 1024 subframes. The BA starting sequence control field is used to decide which BAR the frame responds to.

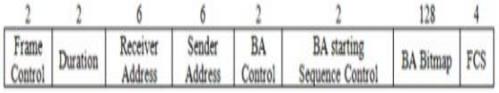


Fig. 6: Block ACK in FA-MPDU scheme

IV. Performance Evaluation

This section contains the results of the FA-MPDU performance. The results of previous studies on normal A-MPDU and DCF are compared. The simulation tests are used to determine the efficiency of FA-MPDU under different subframe sizes and number of stations. The results are presented using the parameters listed in Tables 1 and 2 for various types of traffic.

TABLE : Simulation Parameters

Parameters	Duration (µs)
T_{DIFS}	34
T_{SIFS}	16
$T_{Phy\ hdr}$	20
CW_{min}	16
Retransmission Limited	4 times

TCP Traffic in FA-MPDU Scheme

Figure 7 shows FA-MPDU throughput in comparison to A-MPDU and DCF subframe sizes. When compared to previous literature schemes, the FA-MPDU shows a better performance. When the BER = 10−4, FA-MPDU achieves a throughput of 45 Mbps with a subframe size of 128 bytes and 47 bytes with a subframe size of 256 bytes. when BER = 10−5, the throughput increases to 75 Mbps for subframe of 128 bytes and 77 Mbps for subframe of 256 bytes. When BER = 10−6 and the subframe size is 128 bytes, the throughput reaches 110 Mbps; when the subframe size is 256 bytes, it reaches 111 Mbps. A-MPDU and DCF hold the queued subframe and retransmit until their frames are successfully received in an RD. When BER = 10−4, 10−5, and 10−6, respectively, the performance gain of the FA-MPDU over the A-MPDU varies depending on the subframe size, varying from 6% to 12%, 6% to 7%, and 10% to 13%. The improvement of the system performance with the FA-MPDU is more noticeable than with the other schemes.

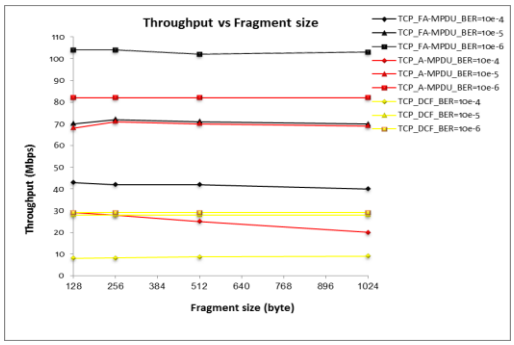


Fig. 7: Throughput vs. Subframe Size

Figure 8 shows the performance throughput of FA-MPDU with 10 to 80 stations. The subframe size is maintained with 256 bytes. All the stations share a common medium. This throughput is achieved by the entire system rather than by a single station. The throughput of FA-MPDU achieves a higher performance compared with the previous schemes. The throughput of FA-MPDU achieves 30, 100, and 130 Mbps at 10 stations when BER = 10^{-4}, 10^{-5}, and 10^{-6}, respectively.

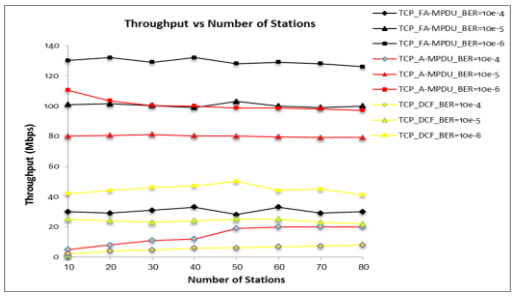


Fig. 8: Throughput vs. Number of Stations

Figure 9 illustrates the FA-MPDU scheme delay in performance and compares it to different subframe sizes. There are 50 stations in total, each with a different BER. Tables 1 and 2 list the remaining parameters. FA-MPDU has a faster response time than the other schemes. An increase in the size of the subframe corresponds to an increase in delay. As soon as FA-MPDU receives the block ACK, it removes the correctly received subframes from the Sq and Rq. The DCF scheme, on the other hand, keeps the subframe in the queue and keeps retransmitting until the complete frame is successfully transmitted. The DCF was able to achieve a long delay. Therefore, transmission times for small subframe sizes are short. The maximum delay of FA-MPDU and A-MPDU when BER = 10−6 is achieved when the subframe size is 1024 bytes with 0.0265 and 0.0350 s, respectively.

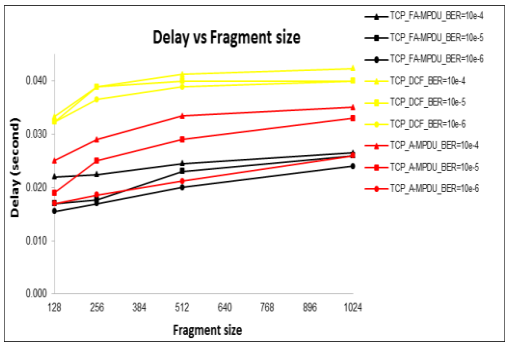


Fig. 9: Delay vs. Fragment Size

Figure 10 depicts the FA-MPDU performance delay and the comparison of literature schemes at 256 bytes for various number of stations and subframe sizes. Because of its subframe-based retransmissions, the FA-MPDU scheme has the shortest delay, as valid subframes are not retransmitted. When BER = 10^{-6}, the FA-MPDU delay ranges from approximately 0.03 s for a network of 10 contending stations to approximately 0.053 s for a network of 90 stations. The delay is accelerated when the number of stations increases, resulting in a superframe retransmission period for literature schemes.

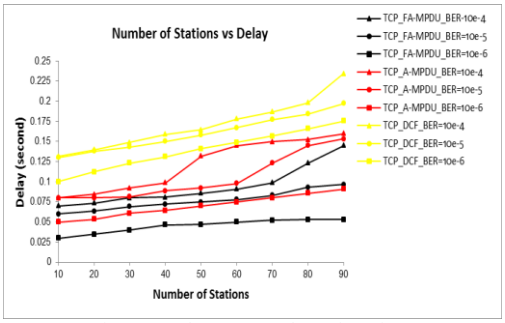


Fig. 10: Delay vs. Number of Stations

V. CONCLUSION

The efficiency of the FA-MPDU performance in terms of throughput and delay was evaluated in this paper. The results are then compared with the previous schemes. In addition, the efficiency of the MAC performance as compared with the standard DCF was also presented. To recognize the A-MPDU transmission scheme, the A-MPDU transmission was elaborated and integrated with the fragmentation method. In noisy channels, the FA-MPDU uses different subframe sizes. By retransmitting only the corrupted subframes rather than the whole frame, the FA-MPDU saves airtime and reduces retransmission overhead. Based on subframe size and the number of stations with aggregation unit, simulation results showed substantial improvements in throughput and delay.

REFRENCES

[1] I. L. M. S. Committee, "IEEE Standard for Information technology—Telecommunication and information exchange between systems—Local and metropolitan area networks—Specific requirements Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment1: Radio Resource Measurement of Wireless LANs," *IEEE, http://standards. ieee. org/getieee802/download/802.11 n-2009. pdf*, 2009.

[2] A. A. Milad, S. M. Lafi, and M. A. Algaet, "Piggyback Scheme over TCP in Very High Speed Wireless LANs."

[3] A. A. Milad, B. M. Noh, Z. Azri, A. S. Shibghatullah, and M. A. Algaet, "Design a novel reverse direction transmission using piggyback and piggyback with block ACK to improving the performance of MAC layer based on very high speed wireless lans." pp. 263-266.

[4] A. Saif, M. Othman, S. Subramaniam, and A. H. NA, "Frame aggregation in wireless Networks: Techniques and Issues," *IETE Technical Review,* vol. 28, no. 4, pp. 336, 2011.

[5] T. Li, Q. Ni, D. Malone, D. Leith, Y. Xiao, and T. Turletti, "Aggregation with fragment retransmission for very high-speed WLANs," *IEEE/ACM Transactions on Networking (TON),* vol. 17, no. 2, pp. 591-604, 2009.

[6] A. Saif, and M. Othman, "SRA-MSDU: Enhanced A-MSDU frame aggregation with selective retransmission in 802.11 n wireless networks," *Journal of Network and Computer Applications*, 2013.

[7] A. Saif, and M. Othman, "A Reliable A-MSDU Frame Aggregation Scheme in 802.11 n Wireless Networks," *Procedia Computer Science,* vol. 21, pp. 191-198, 2013.

[8] J. Barcelo, A. L. Toledo, C. Cano, and M. Oliver, "Fairness and convergence of CSMA with enhanced collision avoidance (ECA)." pp. 1-6.

[9] L. Sanabria-Russo, A. Faridi, B. Bellalta, J. Barcelo, and M. Oliver, "Future evolution of CSMA protocols for the IEEE 802.11 standard." pp. 1274-1279.

[10] E. Perahia, and R. Stacey, *Next generation wireless LANs*: Cambridge University Press Cambridge, 2008.

[11] A. A. Milad, Z. A. B. M. Noh, A. S. Shibghatullah, S. Sahib, R. Ahmad, and M. A. Algaet, "Transmission Control Protocol Performance Comparison Using Piggyback Scheme In Wlans," *Journal of Computer Science,* vol. 9, no. 8, pp. 967, 2013.

[12] R. P. F. Hoefel, "on ieee edca 802.11 n." pp. 478-482.

[13] J. Barcelo, B. Bellalta, A. Sfairopoulou, C. Cano, and M. Oliver, "CSMA with enhanced collision avoidance: A performance assessment." pp. 1-5.

[14] A. Saif, M. Othman, S. Subramaniam, and N. A. W. A. Hamid, "An enhanced a-msdu frame aggregation scheme for 802.11 n wireless networks," *Wireless Personal Communications,* vol. 66, no. 4, pp. 683-706, 2012.

[14] A. Ahmad Milad, M. Noh, Z. Azri, A. S. Shibghatullah, A. Mustapha, and M. A. Algaet, "Reverse direction transmission using single data frame and multi data frames to improve the performance of MAC layer based on IEEE 802.11 n," *Science International-Lahore*, pp. 1861-1864, 2014.

[15] A. A. Milad, M. Noh, Z. Azri, A. S. Shibghatullah, and M. A. Algaet, "Reverse Direction Transmission in Wireless Networks: Review," *Middle-East Journal of Scientific Research*, pp. 767-778, 2013.

[16] N. Hajlaoui, I. Jabri, M. Taieb, and M. Benjemaa, "A frame aggregation scheduler for QoS-sensitive applications in IEEE 802.11 n WLANs." pp. 221-226.

[17] B. Drilo, and L. Flatz, "Comparison of IEEE 802.11g optional standard elements in WLAN hotspot scenario." pp. 147-151.

[18] A. A. Milad, Z. A. B. M. Noh, A. S. Shibghatullah, M. A. Algaet, and A. Mustapha, "Design a New Bidirectional Transmission Protocol to Improve the Performance of MAC Layer Based on Very High Speed WLANs," *Journal of Computer Science,* vol. 11, no. 5, pp. 707, 2015.

[19] E. Perahia, "IEEE 802.11 n development: history, process, and technology," *Communications Magazine, IEEE,* vol. 46, no. 7, pp. 48-55, 2008.

[20] E. Charfi, L. Chaari, and L. Kamoun, "Fairness of the IEEE 802.11 n aggregation scheme for real time application in unsaturated condition." pp. 1-8.

[21] E. Charfi, L. Chaari, and L. Kamoun, "Analytical analysis of applying aggregation with fragment retransmission on IEEE 802.11 e EDCA network in saturated conditions." pp. 1-9.

Performance Analysis of the TCCP Security Protocols

1st Said Naser Said Kamil
Faculty of Science-Khoms
Elmergib University
Khoms, Libya
said.kamil@elmergib.edu.ly

2nd Nigel Thomas
School of Computing
Newcastle University
Newcastle upon Tyne, UK
nigel.thomas@newcastle.ac.uk

***Abstract*—The Trusted Cloud Computing Platform (TCCP) offers a secure execution environment at the Infrastructure as a Service (IaaS) level for customers' VMs. This paper presents an approach for modelling and analysing two security protocols (Node Registration and VM Launch) that form part of the TCCP. We investigate the scalability of these protocols by means of the Markovian process algebra PEPA.**

***Index Terms*—Performance modelling, TCCP security protocol, Node Registration, VM Launch, Cloud Computing, PEPA**

I. Introduction

Cloud computing offers several advantages, for instance, pay-per-use cost, high performance and scalability. However, there are some security aspects that affect and limit the adoption of cloud computing [1], [2], [3], [4]. Service models that are provided by cloud computing, i.e. Software as a Service (SaaS), Platform as a service (PaaS) and Infrastructure as a Service (IaaS) can be exposed to different security breaches at each level [5]. Therefore, understanding the relationships and dependencies between these models of services is substantial to recognize the associated security concerns.

Trusted Cloud Computing Platform (TCCP) system has been presented by [6], that proposes a secure execution environment for the customers of cloud computing using a set of security protocols that work at the infrastructure level. The TCCP guarantees confidentiality customers' VM. Also, it allows customers to determine whether the provided service is secure before the launch of the VMs. Essentially, the proposed TCCP consists of three protocols: Node Registration, VM Launch and Migration. The TCCP offers a number of security improvements. However, in terms of scalability the TCCP is exposed to a major limitation for the reason that all processes are directed and controlled by the Trusted Coordinator (TC), which eventually leads to a bottleneck [7], [8]. Consequently, the performance and service availability will be considerably affected.

In this paper, we present new models of the TCCP protocols for Node registration and VM launch. This work, therefore, aims to further investigate the scalability issues of the TCCP protocols that have been discussed in [7] and [8]. Precisely, we will present a methodology that facilitates greater scalability in the analysis of these protocols, which will allows us to derive performance predictions and approximations of the behaviour of the Trusted Coordinator TC and the Cloud Manager CM under high loads. The paper is organised as follows. The study starts with reviewing some related work. Then a brief overview about the TCCP protocols (Node Registration and the VM Launch) is presented. This is followed by the PEPA model of the Node Registration and its experimental results. After that, a description about the VM Launch PEPA model is given as well as to present the results of the VM Launch PEPA model. Finally, the paper ends with some conclusions and highlights some future works.

II. Related Work

Several approaches have addressed the security and privacy at the level of Infrastructure as a Service of the TCCP system, for example [9], [10]. Additionally, they have presented slightly different designs of the TCCP in order to tackle scalability issues. Han-zhang and Liu-sheng [9] have presented a study to improve the trusted cloud computing platform. They have used Direct Anonymous Attestation (DAA) and trusted third party privacy Certification Authority (CA) schemes to ensure the privacy of cloud providers. Several changes have been made on the original TCCP design; for instance, the introducing of a Trusted Cloud Manager (TCM) which can host many Trusted Coordinators (TC) and nodes within one zone. Whereas they can link to each other through a Cluster Controller (CC).

Although, a considerable number of changes that have been made on the original design of the TCCP to improve the security and the availability, however, there is a significant security weakness in one of the important roles of the TCCP. Whilst, the Cloud Manager (CM) should have no privilege into the external trusted entity that hosts the TC, in the way that is proposed in [9] the TCM controls the communications between a TC and the CC. Furthermore, the connections between the newly invented entities will increase the communications costs and thus affecting the performance significantly.

A Distributed Trusted Cloud Computing Platform (DTCCP) has introduced by [10], which extends the TCCP. The DTCCP proposes that managing all trusted nodes should be distributed to multiple entities rather than being handled by a single

entity. Thus, each entity can manage a cluster of trusted nodes. That allows the system to overcome the undesired bottleneck. On the other hand, the proposed DTCCP relies mainly on the original design of the TCCP hence inherits some of its limitations. Furthermore, the proposed DTCCP has no evidence for any implementation or verification as yet.

III. Trusted Cloud Computing Platform TCCP

A design of the Trusted Cloud Computing Platform (TCCP) has been proposed by [6] based on trusted computing technologies. Whereas a secure execution environment is provided to the cloud service providers at the IaaS level of services. Several techniques are leveraged by the TCCP, for instance, Trusted Platform Module (TPM) [11] and the Trusted Virtual Machine Monitor (TVMM). To provide more security for the VMs of the customers, only the Trusted Coordinator (TC) manages the trusted nodes. Also, the TC is hosted and maintained by a third party External Trusted Entity (ETE) where the cloud manager (CM) has no privileges in the ETE. For more details about the TCCP security protocols, including the Node Registration and the VM Launch. We refer the interested reader to [6].

A. Node Registration Protocol

As stated by [6], for a node to become trusted it must comply with the protocol where the encryption and decryption of the private and public keys are used. Additionally, attestations are used to guarantees authenticity between Node and TC.

B. Virtual Machine Launch Protocol

After applying the Node Registration protocol a node will be registered as a trusted node. Accordingly, the trusted nodes exclusively will be qualified for hosting and launching the VMs of the customers [6]. The VM Launch protocol uses the exchange of the encrypted and decrypted private and public keys between User, CM, Node and TC for the verifications and to fulfil the requirements. Where the TCCP allows users to decide whether to use the VM or not.

IV. Node Registration PEPA Model

In the PEPA model presented by [8] there was an issue when the number of nodes was scaled up. Whereas the performance of the *TC* is significantly affected by the increase of the number of nodes. As a consequence, the throughput of the studied model was limited. Therefore, a new model will be created and analysed using the PEPA Eclipse Plug-in tool [12], with the aim to address the scalability of the PEPA model of [8].

In the following Node Registration model a *Node*, sends a registration request and proceed through the protocol processes to be a trusted *Node*. In the case, it is trusted the model return to the beginning to make a new request. Accordingly, at any time there are N number of $Nodes$ in the model communicating with the TC. In effect, the model is a closed queuing network.

$$
\begin{aligned}
Node &\overset{def}{=} (sendNonce\_N, r_1).Node_0 \\
Node_0 &\overset{def}{=} (cpu, c).Node_1 \\
Node_1 &\overset{def}{=} (cpu, c).Node_2 \\
Node_2 &\overset{def}{=} (isTrusted, p_1 * r_7).Node_3 \\
&+ (untrusted, (1-p_1) * r_7).Node \\
Node_3 &\overset{def}{=} (sendNonce\_TC, r_{11}).Node_4 \\
Node_4 &\overset{def}{=} (generateKeypairTprivate_N\ Tpublic\_N \\
&\quad , r_5).Node_5 \\
Node_5 &\overset{def}{=} (sendPublicKey\_N, r_6).Node_6 \\
Node_6 &\overset{def}{=} (attestSuccessfully, p_2 * r_8).Node_7 \\
&+ (unsuccessful, (1-p_2) * r_8).Node \\
Node_7 &\overset{def}{=} (cpu, c).Node_8 \\
Node_8 &\overset{def}{=} (sendAccepted, r_{10}).Node_9 \\
Node_9 &\overset{def}{=} (end, r_{14}).Node \\
TC &\overset{def}{=} (sendNonce\_N, r_{100}).TC \\
&+ (cpu, c).TC + (isTrusted, r_{100}).TC \\
&+ (untrusted, r_{100}).TC \\
&+ (sendNonce\_TC, r_{100}).TC \\
&+ (sendPublicKey\_N, r_{100}).TC \\
&+ (attestSuccessfully, r_{100}).TC \\
&+ (unsuccessful, r_{100}).TC \\
&+ (sendAccepted, r_{100}).TC \\
System &\overset{def}{=} Node[N] \underset{\mathcal{L}}{\bowtie} TC
\end{aligned}
$$

The model employs only two components (*Node* and *TC*). The $Node$ component has a sequence of all actions proposed by Santos *et al* [6]. $Node$ has been used to preserve the sequence of all actions. In contrast, the TC component contains only the actions which are performed by a Trusted Coordinator in parallel with the $Node$ component. Those actions which are deemed to require significant action by the Trusted Coordinator namely, *bootstrapML_TC, encryptedPrivateKey_TC* and *addTrustedPublicKey_N* are replaced in the model by a single cpu action, representing the competition for resources at the *TC*. The rates of those actions (i.e. r_2, r_3 and r_9) are used to calculate the rate of the cpu action c. The value of the rate c of the cpu action is calculated as:

$$c = \frac{3}{\frac{1}{r2} + \frac{1}{r3} + \frac{1}{r9}} \tag{1}$$

Where r_2, r_3 and r_9 are the rates of *bootstrapML_TC, encryptedPrivateKey_TC* and *addTrustedPublicKey_N* respectively. This rate represents the average duration rate of those actions. Therefore, the cpu action will be used to represent the contention on resources in the TC. The rest of the TC actions have been kept in the model for completeness' purpose, even though they have little significance on its performance. The rate r_{100} is used as rate of all TC component actions except cpu action. The main purpose of using r_{100} in the model is to allow the independent actions to proceed without any waiting time. So, it means there is a coordination, but the actual actions happen almost independently.

The cpu action is used because PEPA does not easily allow us to directly limit the rates across multiple action

types with a single bound. Hence, we model a single action (cpu) and limit the total rate of this action, however it is in effect a combination of all the actions of which it is comprised (i.e. *bootstrapML_TC, encryptedPrivateKey_TC* and *addTrustedPublicKey_N*). So, cpu cannot run faster than all the comprising actions (bounded capacity).

The system equation gives the cooperation between N instances of *Node* and the *TC* over the set *L*. Where $N = (1000\ to\ 20000)$ and the cooperation set $L = \{$*sendNonce_N, cpu, isTrusted, untrusted, sendNonce_TC, sendPublicKey_N, attestSuccessfully, unsuccessful, sendAccepted*$\}$. The rates used in this model are shown in the Table I.

TABLE I: Node Registration PEPA model rates

Rate	Value	Rate	Value	Rate	Value	Rate	Value
r_1	1	r_6	0.75	r_{10}	0.73	r_{14}	0.0001
r_2	0.4	r_7	0.35	r_{11}	0.46	p_1	0.9
r_3	0.7	r_8	0.65	r_{12}	0.2	p_2	0.9
r_5	0.55	r_9	0.65	r_{13}	0.01	r_{100}	1000
c	$3/(1/r_2 + 1/r_3 + 1/r_9)$						

A. Experiments and Results

1) Model scalability (Large scale): The model scalability will be examined in this section, where the number of $Nodes$ have been varied from 1000 to 20000, also, the TC and refresh rate are used with the values $TC = 1, 2$ and $r_{14} = 0.0001$. As can be seen, Figure 1 illustrates the throughput of the cpu action (where it is supposed that there is a contention on the resources), and the $sendAccepted$ action that represents the successfulness of the node registration request. It is obvious that throughput of cpu action saturates at some point between (1000 and 5000) nodes, whereby after 5000 it is just shown as a flat line. By replicating the TC instances ($TC = 2$) the system throughput is doubled in both actions (cpu and $sendAccepted$). It is worth noting that the $sendAccepted$ throughput reaches its peak at 10000 nodes with $TC = 1$ and then it starts decreasing gradually as the number of Nodes is increased. This means that $Nodes$ start queuing waiting for registration because the system is saturated. Furthermore, there is a possibility that there are more $Nodes$ wanting to do a specific action than the capacity of the system. Therefore, a further investigation is required where the system will be examined with a higher value of r_{100} and see how this will affect the throughput drop of the $sendAccepted$ action.

2) Saturation Point Exploration: The motivation of this set of experiments is to explore further the saturation point in the previous experiments that shown in Figure 1, which lies somewhere between ($Node = 1000$ and $Node = 5000$). Model throughput is shown in Figure 2, where the number of nodes varied from 1000 to 5000. Obviously, the analysis showing that the saturation point for the cpu throughput with one TC is at $Node = 2000$ and then with two TCs is at $Node = 5000$. While, with one TC the throughput of $sendAccepted$ action is saturated at 2000 $Nodes$, replicating the TC will make the throughput of the same action increase steadily with the rise in the number of $Nodes$ as illustrated in Figure 2.

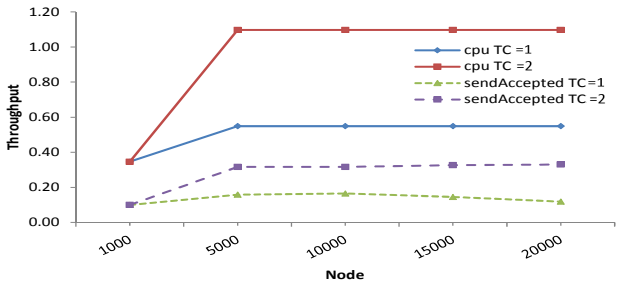


Fig. 1: Throughput of Node Registration model (*cpu & sendAccepted*) actions *N=1000 to 20000, TC=1,2,* r_{100}*=1000*

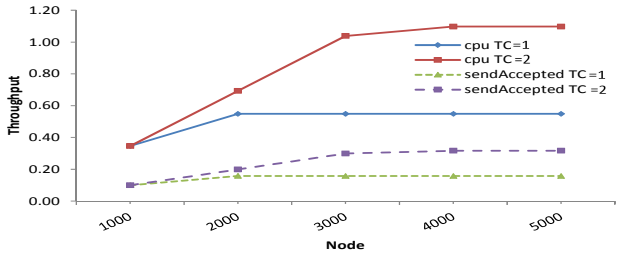


Fig. 2: Throughput of Node Registration model (*cpu & sendAccepted*) actions *N=1000 to 5000, TC=1,2,* r_{100}*=1000*

V. VM Launch PEPA Model

In the previous experiments on the VM Launch PEPA model [7] the *TC* becomes the bottleneck as the number of users is increased. That is because the security protocol centralises all processes to be managed through the *TC*. Therefore, the second motivation of this study is to explore the capacity of the system under question. In other words, a further investigation will take place with the aim of exploring the performance limitations of the previous VM Launch model [7]. To achieve this goal, a new PEPA model will be created, which will simulate the TCCP (VM Launch) security protocol.

A new component, *Node* has been specified that will be used to control the behaviour of the other components engaging in the VM Launch protocol. Thus *Node* has a sequence of all actions and will cooperate in parallel with the components *User, CM* and *TC*. In addition, once a User triggers a request the components (*Node, User, CM* and *TC*) simultaneously can do any of their internal or cooperated actions as long as the sequence of the actions is controlled by the *Node* component. Furthermore, it has been used the *cpu* action instead of the following actions (*sendMessageTo_TC_EncrecptedBy_TKp_N, searchInDatabaseFor_TKP_N, decrypte_Kvm and sendDecrypted_Kvm_To_N*). In fact, it is our assumption that these actions are supposed to use the *TC* extensively, hence, *cpu* action will allow limiting the rate across multiple actions. Consequently, allowing to control the rate of those actions.

The system equation showing the cooperation between the model components (*Node, User, CM* and *TC*) over the set L and there is another cooperation between *User*

and *CM* over the set M. Where N_1 =(100 to 5000) , $N_2 = 15$ and $N_3 = 10$ in the part one of the first set of experiments and then varied for larger values as illustrated in the following sections. Also, the set L ={*generate_Kvm, sendAlphaHashAlphaEncreptedBy_Kvm, key_KvmEncreptedBy_TKP_TC, receiveRequest, designate_N, forwardRequestTo_N, cpu, isTrusted_N, untrusted_N, request_isDenied, send_N_id_To_User_EncreptedBy_Kvm, forward_N_id_To_User, useVM*}. Additionally, the set $M = \{$ *forward_N_id_To_User* $\}$.

$$
\begin{aligned}
Node &\stackrel{def}{=} (generate\_Kvm, r_1).Node_0 \\
Node_0 &\stackrel{def}{=} (sendAlphaHashAlphaEncreptedBy\_Kvm, r_2).Node_1 \\
Node_1 &\stackrel{def}{=} (key\_KvmEncreptedBy\_TKP\_TC, r_3).Node_2 \\
Node_2 &\stackrel{def}{=} (receiveRequest, r_4).Node_3 \\
Node_3 &\stackrel{def}{=} (designate\_N, r_5).Node_4 \\
Node_4 &\stackrel{def}{=} (forwardRequestTo\_N, r_6).Node_5 \\
Node_5 &\stackrel{def}{=} (cpu, r_7).Node_6 \\
Node_6 &\stackrel{def}{=} (cpu, r_8).Node_7 \\
Node_7 &\stackrel{def}{=} (isTrusted\_N, p * r_9).Node_8 \\
&+ (untrusted\_N, (1-p) * r_9).Node_{10} \\
Node_8 &\stackrel{def}{=} (cpu, r_{10}).Node_9 \\
Node_9 &\stackrel{def}{=} (cpu, r_{11}).Node_{11} \\
Node_{10} &\stackrel{def}{=} (request\_isDenied, r_{12}).Node \\
Node_{11} &\stackrel{def}{=} (decrypt\_alpha\_bootVM, r_{14}).Node_{12} \\
Node_{12} &\stackrel{def}{=} (send\_N\_id\_To\_User\_EncreptedBy\_Kvm, r_{16}).Node_{13} \\
Node_{13} &\stackrel{def}{=} (forward\_N\_id\_To\_User, r_{17}).Node_{14} \\
Node_{14} &\stackrel{def}{=} (useVM, r_{18}).Node \\
User &\stackrel{def}{=} (generate\_Kvm, r_1).User \\
&+ (sendAlphaHashAlphaEncreptedBy\_Kvm, r_2).User \\
&+ (key\_KvmEncreptedBy\_TKP\_TC, r_3).User \\
&+ (forward\_N\_id\_To\_User, r_{17}).User \\
&+ (useVM, r_{18}).User \\
CM &\stackrel{def}{=} (receiveRequest, r_4).CM \\
&+ (designate\_N, r_5).CM \\
&+ (forwardRequestTo\_N, r_6).CM \\
&+ (send\_N\_id\_To\_User\_EncreptedBy\_Kvm, r_{16}).CM \\
&+ (forward\_N\_id\_To\_User, r_{17}).CM \\
TC &\stackrel{def}{=} (cpu, c).TC + (isTrusted\_N, p * r_9).TC \\
&+ (untrusted\_N, (1-p) * r_9).TC \\
&+ (request\_isDenied, r_{12}).TC \\
System &\stackrel{def}{=} Node[N_1] \underset{\mathcal{L}}{\bowtie} ((User[N_1] \underset{\mathcal{M}}{\bowtie} CM[N_2]) \\
&\| \; TC[N_3])
\end{aligned}
$$

The rates that used in this model are exactly the same to those rates that have been used in [7] and are shown in the Table II. As well as to use c rate, which is experimental rate and represents the rate of the *cpu* action.

TABLE II: The rates of the VM Launch Protocol PEPA model

Rate	Value	Rate	Value	Rate	Value	Rate	Value
r_1	10.0	r_5	6.0	r_9	8.0	r_{13}	1.0
r_2	0.4	r_6	8.0	r_{10}	6.9	r_{14}	3.5
r_3	0.7	r_7	6.5	r_{11}	9.0	r_{16}	6.8
r_4	3.9	r_8	7.0	r_{12}	8.6	r_{17}	7.2
r_{18}	0.05	p	0.8	c	20		

A. Experiments and Results

The model scalability will be examined in this section, whereas two sets of experiments will be presented. Whilst, the first set will replicate the resources (*CM* and *TC*), the second set will increase the rate of actions instead of replicating the servers. Furthermore, in this model we have assumed that the number of nodes is equal to the number of users, this to avoid the case where the *User* component become a bottleneck. It is noteworthy that only two most significant actions will be presented throughout our results, which is the *cpu* action that presents the maximum throughput of the model; and the *useVM* action, which is demonstrating the successfulness of a launch request.

1) Scalability by Replicating Servers (CM and TC): Initially, it has been looked at the scalability of the model where the number of *Nodes* and *Users*= 1000, *CM*= 15 and *TC*= 10, i.e. this is the balance point. As can be seen in Figure 3, the throughput of the model is increased as we increase the number of nodes up to *Node*= 1000, which is the saturation point. After this point, the throughput is shown as a flat line because the system is saturated, which means not enough capacity at the server side. That is to say, increasing the number of nodes larger than 1000 in the case that shown in Figure 3, will only make any coming request to just join the waiting queue. Thus, waiting for a free resource to be processed. Moreover, the maximum throughput for the *TC* can be calculated as the number of *CPUs* (i.e. in *TC* instances) multiplied by the rate of the *cpu* action (c). Means it is used all the time (the utilisation is 100%), as a result, it is saturated.

On the other hand, the scalability of the model can be increased substantially by only increasing the number of *CMs* to be 72 and the *TCs* to be 50 instances as illustrated in Figure 4. Additionally, unlike the results are shown in Figure 3, the model now able to handle 5000 requests simultaneously. But, also, again the model is saturated at 5000 because it reaches its maximum throughput and will not able to process any coming requests while waiting for a resource to become free. So, this means, the model is able to scale up as we increase the number of resources instances, which is allowed to predict the saturation point of a system for any given load (i.e. a number of nodes) and then can predict the most appropriate number of resources that can handle this load.

Latency is the second metric that has been used in our experiments for the evaluation of the model performance. Where, it has been calculated as follows:

$$Latency = \frac{1}{rate_a} + \frac{Max(0, (Pop - N + 1))}{N * rate_a} \qquad (2)$$

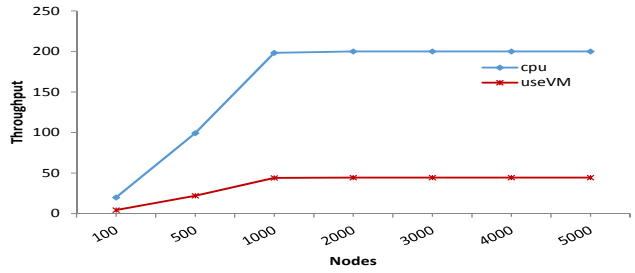


Fig. 3: Throughput of VM Launch, *CM*=15 and *TC*=10

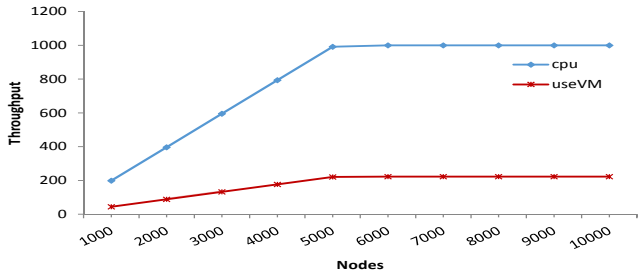


Fig. 4: Throughput of VM Launch, *CM*= 72 and *TC*=50

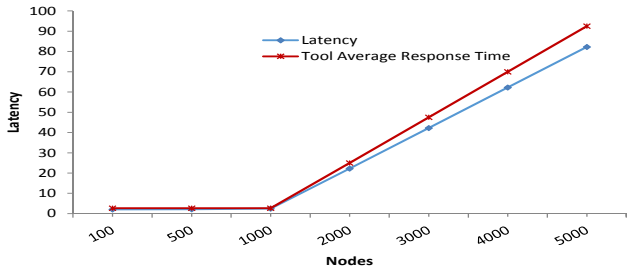


Fig. 5: Latency of VM Launch, where *CM*=15 and *TC*=10

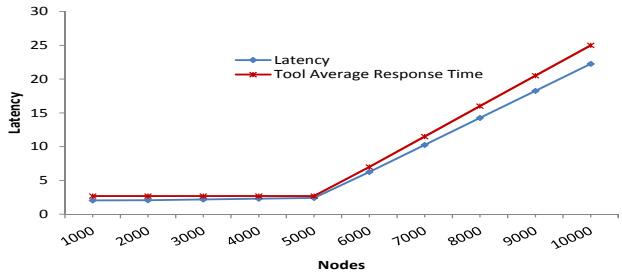


Fig. 6: Latency of VM Launch, where *CM*=72 and *TC*=50

Where *N* is representing the number of servers that cooperate in a specific action which would relate to *CM* or *TC*. Also, it has been validated by comparing our calculation of the latency with the Average Response Time that given by the PEPA Eclipse Plug-in tool. Figure 5 and Figure 6 respectively display the latency of the VM Launch model for the relatively small and large scale systems. It is obvious that, the latency of the model in both figures is shown as a flat line up to the *Node*= 1000 in Figure 5 and then up to the *Node*= 5000 in Figure 6. That is because there are enough servers in both cases respectively. Then it start increases linearly in both figures as the load becomes higher than 1000 nodes in Figure 5 and greater than 5000 nodes in Figure 6. Noteworthy, both calculations (calculated latency and the tool Average Response Time) are very consistent in their prediction up to the saturation point where *Node*= 1000 in Figure 5 and *Node*= 5000 in Figure 6. However, there is a small deviation in both figures, specifically, after the prediction of the saturation point. Also, Figure 6 showing the model latency is decreased significantly even with a large number of nodes in comparison with Figure 5. That is because in Figure 6 the number of servers is increased (*CM*= 72 and *TC*= 50), thus the system gives better performance.

2) Scalability by Changing Rates: Alternatively, in the following set of experiments, it has been increased the rates of actions instead of raising the number of servers, i.e. *CM* and *TC*. Indeed, it is another way of considering the VM Launch security protocol. In practise, it is not desired to have, for example, 72 *CMs* and 50 *TCs* instances, and also, having replicated servers will increase the vulnerability of the platform. Nevertheless, this kind of model allowing us to provide an approximation of the performance of such systems at an abstract level. That is to say, having, for instance, one *CM* that 72 times faster, will be cost effective and offer an enhanced performance. In fact, increasing the number of instances in a model is effectively increasing the rates, where it will be equals to the number of instances multiplied by those rates. Table III represents the rates that have been used in the model in the second set of experiments. Where it has been considered model scalability by changing the actions rates and all these rates are experimental rates.

TABLE III: VM Lunch rates of actions are increased

Rate	Value	Rate	Value	Rate	Value
r_1	300	r_7	325	r_{14}	175
r_2	200	r_8	350	r_{16}	489.6
r_3	280	r_9	400	r_{17}	518.4
r_4	320	r_{10}	345	r_{18}	0.05
r_5	432	r_{11}	450	c =	1125
r_6	576	r_{12}	430		

Although, Figure 7 exhibiting the model results with different specifications in the model, the results are very close to those illustrated in Figure 4. However, it is worth noting that the system now is able to handle 5000 nodes concurrently with these fast rates and only one *CM* and one *TC*. So, once it becomes saturated we can replicate the number of servers in the model. Hence, increasing the number of instances in this case of (Figure 7) will significantly increase the model scalability. Accordingly, it will be able to scale up to deal with a larger number of nodes.

In comparison with the latency that shown in Figure 6, Figure 8 gives better results where the latency of the model is decreased. Also, the latency is shown as a flat line, because there is no competition for resources. Moreover, the comparison of the calculated latency and the tool Average Response Time that displayed in Figure 8 is very consistent up to the point where *Node*=5000 (i.e. have enough resources), then start increasing linearly as the queuing time rises.

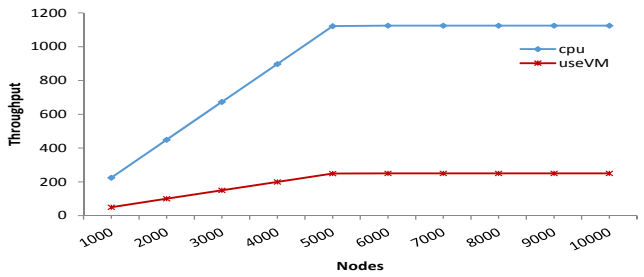


Fig. 7: Throughput of VM Launch using rates of Table III

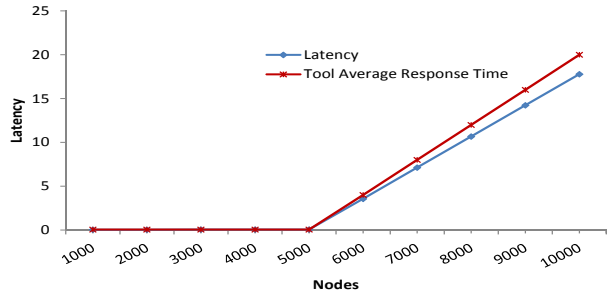


Fig. 8: Latency of VM Launch by using rates of Table III

B. Discussion

In practise for any given resources there is a capacity limit, nevertheless, with this model, we are able to predict the saturation point for a given number of nodes and users. Accordingly, at the saturation point we able to determine whether the bottleneck will become on the *CM* side or on the *TC* side. Both cases that have been illustrated in the above figures agree to some extent. But, the numbers are slightly different as a result of how the ODEs analysis dose it. Whereas, the ODEs gives more precise outcomes with enormous components' amount. So, if the capacity exceeded, then it is going to be saturated. Thus, result in a steadily increasing latency as displayed in Figure 5, Figure 6 and Figure 8. It is noticeable that, is notionally and numerically slightly different approximation between Figure 4 and Figure 7 with the same number of nodes. But, that is because in the latter figure we have used fast rates, which allow to scale up with less number of servers.

VI. Conclusion

We have presented approaches for analysing the performance of the Trusted Cloud Computing Platform TCCP system, specifically the Node Registration and VM Launch protocols. This work is motivated by investigating the scalability issues of the Node Registration and the VM Launch protocols that have been explored in [7] and [8] that forms part of a larger study into scalable modelling of protocols for cloud computing.

Experiments with the Node Registration protocol model have shown that it is able to scale up with the increased loads. As a result, replicating the *TC* instances will allow the model significantly to scale up with the increase of the number of nodes. Hence, this model is able to cope with the performance limitation of the previous model [8]. In practise, there are other reasons that might limit the *TC* performance, for instance, network bandwidth, which is not considered in this research.

The VM Launch model that illustrated in section V, is not only able to scale up with large loads, but also, can be used to predict the saturation point for the given load. We have explored two different cases. First, by scaling the number of servers of both *CM* and *TC*. Second, by scaling the model through the rise of the rates. Obviously, having changed the number of servers in the model gives a fine granularity. Which means that effectively will get what is the capacity needed at the *TC* and the *CM*, in order to get a scalable result. Consequently, the model is able to provide the scalability while maintaining the confidentiality and integrity. Clearly, it is more beneficial to validate the obtained results against a real augmentation. Whereas, one of the objectives of modelling training study would be to perform validations of all created models. But in this case, we do not have an implementation of the TCCP system. So, the future work is to consider other similar types of protocol, where is an implementation on which we can do the validation against the model.

References

[1] A. Celesti, M. Fazio, M. Villari, A. Puliafito, and D. Mulfari. Remote and deep attestations to mitigate threats in Cloud Mash-Up services. In 2013 World Congress on Computer and Information Technology (WCCIT), pages 16, June 2013.

[2] Keiko Hashizume, David G. Rosado, Eduardo Fernndez-Medina, and Eduardo B. Fernandez. An analysis of security issues for cloud computing. Journal of Internet Services and Applications, 4(1):113, 2013..

[3] David G. Rosado, Rafael Gmez, Daniel Mellado, and Eduardo FerndezMedina. Security Analysis in the Migration to Cloud Environments. Future Internet, 4(2):469, 2012.

[4] Joanna Rutkowska. Introducing stealth malware taxonomy. COSEINC Advanced Malware Labs, pages 19, 2006.

[5] Cloud Security Alliance. Security guidance for critical areas of focus in Cloud Computing V3.0. https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf, 2011. Online; accessed 05 May 2015.

[6] Nuno Santos, Krishna P. Gummadi, and Rodrigo Rodrigues. Towards Trusted Cloud Computing. In Proceedings of the 2009 Conference on Hot Topics in Cloud Computing, HotCloud09, Berkeley, CA, USA,USENIX Association. 2009.

[7] Said Kamil and Nigel Thomas. A Performance Model of the Trusted Cloud Computing Platform VM Launch Protocol. In Proceedings of the 9th EAI International Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS15, pages 216219. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016.

[8] Said Naser Said Kamil and Nigel Thomas. Performance analysis of the trusted cloud computing platform. In 31st UK Performance Engineering Workshop. School of Computing University of Leeds, 2015.

[9] Wang Han-zhang and Huang Liu-sheng. An improved trusted cloud computing platform model based on DAA and privacy CA scheme. In 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), volume 13, pages V1333V1339, Oct 2010.

[10] P. Sen, P. Saha, and S. Khatua. A distributed approach towards trusted cloud computing platform. In Applications and Innovations in Mobile Computing (AIMoC), 2015, pages 146151, Feb 2015.

[11] Trusted Computing Group (TCG). Trusted Platform Module (TPM). http://www.trustedcomputinggroup.org/resources/trusted platform module tpm summary, 2008. Online; accessed 15 September 2021.

[12] Mirco Tribastone, Adam Duguid, and Stephen Gilmore. The PEPA Eclipse Plugin. SIGMETRICS Perform. Eval. Rev., 36(4):2833, 2009.

Performance Evaluation of First Hop Redundancy Protocol (FHRPv6) with Routing Protocol OSPFv6

Mahmud Mansour
Network Department
University of Tripoli
Tripoli Libya
mah.mansour@uot.edu.ly

Mohmed K. Al-said
Network Department
University of Tripoli
Tripoli Libya
mohamedsid524@gmail.com

Mohamed Alamin Alqomati
Network Department
University of Tripoli
Tripoli Libya
malqmaty041@gmail.com

***Abstract*—Network availability is a key consideration in disaster planning, but it also has critical impacts on everyday life and work. For organizations, network downtime or sluggishness equates to business downtime, at considerable cost to organizations through inefficiency, lost sales, lack of critical data for decisions, and other harmful effects. First hop redundancy protocols (FHRP) are an essential tool for improving the availability of IP networks. In this paper, we evaluate the three particular protocols of FHRPs, namely the Hot Standby Router Protocol (HSRPv6), Virtual Router Redundancy Protocol (VRRPv3), and Gateway Load Balancing (GLBPv6) using GNS3 tools. The First Hop Redundancy Protocols have been implemented, tested, optimized, and compared to one another in terms of CPU Utilization, Traffic flow, packet loss and convergence time. The comparison indicates which protocol is best in which scenario and which is best among the three protocols.**

Keywords—VRRP, FHRP, HSRP, GLBP

I. INTRODUCTION

In the modern world, most enterprises are becoming more considerate about the network availability and minimization of downtime because the demand for online applications and services has increased rapidly. This resulted in increasing demands for services that provide network availability and minimize network downtime for these businesses.
In today's network, availability became a major issue for corporations and businesses. Each minute of outage could cause a company hundred, if not thousands of dollars. In order to minimize outages, we try to increase the uptime of the network by using redundant links and nodes. Although redundancy is good it is costly too, and there is no single way of achieving optimal availability for network it depends on the customer business needs and how much it can tolerate the downtime of the network [1].

II. AVAILABILITY

Availability refers to the amount of time a network is available to users and is often a critical goal for network design customers. The availability can be expressed as a percent uptime per year, month, week, day, or hour, compared to the total time in that period. Availability is linked to reliability but has a more specific meaning (percent uptime) than reliability. Reliability refers to a variety of issues, including accuracy, error rates, stability, and the amount of time between failures [3].

A. Measuring Availability

Network availability is measured as the percentage of time a system stays fully operational over a period of time, usually over a year. Service providers will typically include a specified level of network availability in a service level agreement (SLA).

Availability is also associated with resiliency, which is a word that is becoming more popular in the networking field. Resiliency means how much stress a network can handle and how quickly the network can rebound from problems including security breaches, natural and unnatural disasters, human error, and catastrophic software or hardware failures. A network that has good resiliency usually has good availability [2].

To calculate a theoretical availability, the network is divided into each dependent unit, such as hardware, software, physical connections, power supplies etc. For most equipment, the manufacturer will supply information on availability expectations, often described as Mean Time Between Failure (MTBF).

For those parts of the network not having this data, such as a power source, statistical data and estimations have to be used. The expected time to repair each part of the network has to be estimated. This is normally referred to as Mean Time to Repair (MTTR). The availability for each unit is calculated by:

$$Availability = \frac{MTBF}{MTBF+MTTR} \quad (1)$$

The total availability for the network is then determined by combining the availabilities of the individual components. Theoretically, the availability of a redundant network should be higher than a serially connected one. However, the time taken to fail-over to the standby device should also be considered in the redundancy calculations [2].

Network redundancy is a procedure that involves including additional instances of network devices and equipment in a network as a way of ensuring network availability in the event that a network device or network path fails. Redundancy can be implemented at layer 2 using spanning tree protocol but this paper looks at redundancy options at the network layer using first hop redundancy protocol.

B. Cost of Network Downtime

Network downtime occurs when this digital network shuts down or becomes unavailable for use. Downtime can be either planned or unplanned.

Many organizations do not fully understand the impact of downtime on their business. Calculating the cost of this impact can be difficult because it requires an understanding of both tangible and intangible losses. Tangible losses are quantifiable, hard costs; they include lost revenue, the cost to recover lost information, disaster recovery, and business continuity costs. Intangible costs include damage to your company's reputation, lost customers, and employee productivity costs. In many ways, the damage associated with intangible costs can have a greater long-term impact on an organization than that of tangible costs.

In 2020, the information Technology intelligence consulting (ITIC) study showed Fig (1) that since 2016 the average cost of downtime that lasts 1 hour has risen by 30%. In summary, 1,000 companies answered the poll questions, and the results were as follow [4].

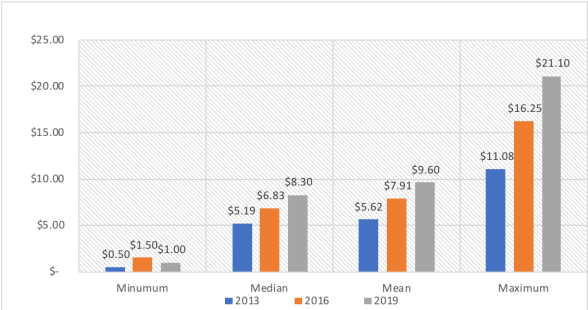


Fig. 1. ITIC Average cost

More than 30% of the enterprises claimed that they spend from $1 to 5 million on 1 hour of downtime. Meanwhile, over $300,000 is the value of 1 hour of downtime for nearly 80% of organizations. Finally, 98% reported that 1 hour of downtime costs them almost $100,000 [21].

III. Related Work

Besides, research [15] conducted by A. Zemtsov. (2019) entitled "Performance Evaluation of First Hop Redundancy Protocols for a Computer Networks of an Industrial Enterprise".

In a previous study [16], M. Mansour, et al (2021) under the title "Performance Analysis and Functionality Comparison of First Hop Redundancy Protocols" investigate the impact of several factors such as Packet loss, CPU utilization, convergence time.

In a previous study [17], M. Mansour (2020) under the title "Performance Evaluation of First Hop Redundancy Protocols" investigate the impact of several factors such as convergence time, CPU utilization, Bandwidth consumption, Traffic flow.

In addition, another paper [18] study by Usman et al (2019) entitled "Performance Analysis and Functionality Comparison of FHRP Protocols" investigate the impact of the bandwidth usage, CPU utilization and convergence time were measured.

Another study [19] by Rahman et al. (2017) titled "Performance Evaluation of First Hop Redundancy Protocols (HSRP, VRRP & GLBP)" where this study was conducted to evaluate the performance of HSRP, VRRP, and GLBP with only one parameter, namely packet loss.

Research [20] conducted by Imelda et al (2020) with the title "Performance Analysis of VRRP, HSRP, and GLBP with EIGRP Routing Protocol" which evaluates the three FHRP protocols, namely VRRP, HSRP, and GLBP and tests using parameters throughput, delay, packet loss, and downtime. But it is using one routing protocol that is EIGRP.

IV. First Hop Redundancy Protocols

First, Hop Redundancy Protocol (FHRP) is a group of protocols that allow a router on a network to automatically take over if a primary default gateway router fails. The devices on a shared network segment are configured with a single default gateway address that points to the router that connects to the rest of the network. The problem comes when this primary router fails, and there is a second router on the segment that is also capable of being the default gateway but end devices don't know about it. Hence, if the first default gateway router fails, the network will terminate [2] [4].

One of the solutions to this problem is First Hop Redundancy Protocols. The three main First Hop Redundancy Protocols are HSRP - VRRP -GLBP [15].

First hop redundancy protocols such as HSRP and VRRP provide default gateway redundancy with one router acting as the active gateway router with one or more other routers held in standby mode. While others like GLBP enables all available gateway routers to load share and be active at the same time [5] [6].

A. Hot Standby Routing Protocol (HSRPv6)

The HSRP is an FHRP designed to allow for transparent failover of the first-hop IP router. HSRP provides high network availability by providing first-hop routing redundancy for IP hosts on Ethernet configured with a default gateway IP address. HSRP is used in a group of routers for selecting an active router and a standby router in a group of router interfaces, the active router is the router of choice for routing packets; the standby router is the router that takes

over when the active router fails or when preset conditions are met [7].

HSRPv6 hosts learn of available IPv6 routers through IPv6 neighbor discovery route advertisement (RA) messages. These are multicast periodically, or may be solicited by hosts. HSRP is designed to provide only a virtual first hop for IPv6 hosts.

An HSRP IPv6 group has a virtual MAC address that is derived from the HSRP group number, and virtual IPv6 link-local address that is, by default, derived from the HSRP virtual MAC address. Periodic RAs are sent for the HSRP virtual IPv6 link-local address when the HSRP group is active. These RAs stop after a final RA is sent when the group leaves the active state [9] [22].

Periodic RAs for the interface link-local address stop after a final RA is sent while at least one virtual IPv6 link-local address is configured on the interface. No restrictions occur for the interface IPv6 link-local address other than that mentioned for the RAs. Other protocols continue to receive and send packets to this address [24].

HSRP uses a priority mechanism to determine which HSRP configured router is to be the default active router. To configure a router as the active-router, you assign it a priority that is higher than the priority of all the other HSRP-configured routers. The default priority is 100, so if you configure just one router to have a higher priority, that router will be the default active router [8].

1- HSRP IPv6 Virtual MAC Address Range

HSRP routers communicate with each other by exchanging HSRP hello packets. These packets are sent to the destination IP multicast address FF02::66 on UDP port 2029 for IPv6 [29]. HSRP IPv6 uses a different virtual MAC address block than does HSRP for IP: 0005.73A0.0000 through 0005.73A0.0FFF (4096 addresses) [22].

2- HSRPv6 Timers

Hello timer: Hello time is the approximate time that routers send in a hello massage to indicate that the peer router is active, and the default value is 3 seconds.
Hold timer: Hold time is the approximate time that standby router will declare that the peer is dead and becomes active, and the default value is 10 seconds.
These timers are tunable and are tuned to obtain minimum convergence and therefore making a network highly available.

B. Virtual Router Redundancy Protocol (VRRPv3)

Virtual Router Redundancy Protocol (VRRP) is an open standard redundancy protocol for establishing a fault-tolerant default gateway. VRRP provides an alternate route path for hosts without changing the IP address or MAC that the host knows [11]. VRRP adds a group of routers that can act as network gateways that enable the traffic to pass through that gateways. Routers in the VRRP group elect a master through the VRRP election mechanism to act as a gateway. Routers in a VRRP group determine their roles by priority [9].

VRRP version 3 (VRRPv3) introduces IPv6 address support for both standard VRRP and VRRP enhanced (VRRP-E) [23].

VRRPv3 implements support for IPv6 addresses for networks using IPv6, and it also supports IPv4 addresses for dual-stack networks configured with VRRP or VRRP-E. VRRPv3 is compliant with RFC 5798. The benefit of implementing VRRPv3 is faster switchover to backup devices than can be achieved using standard IPv6 neighbor discovery mechanisms. With VRRPv3, a backup router can become a master router in a few seconds with less overhead traffic and no interaction with the hosts [11].

1. VRRPV3 Addressing

VRRP communicates with the other configured VRRP routers (within the same group) by sending an advertisement out every second (by default); this advertisement provides these devices with a state of the current master router as well as its priority. VRRP uses the IPv6 multicast address of FF02::12 that used to send hello messages [25].

2. VRRPv3 Timers

VRRP has the best default timer allowing it to converge faster than HSRP or GLBP with their default timers. It uses millisecond timers where absolutely necessary and with careful consideration and testing. millisecond values work only under favorable circumstances. The use of the millisecond timer values is compatible with third party vendors, as long as they also support VRRPv3. You can specify a timer value between 100 milliseconds and 40000 milliseconds [10].

C. GLBP for IPv6

The Gateway Load Balancing Protocol (GLBP) feature provides automatic router backup for IPv6 hosts configured with a single default gateway on an IEEE 802.3 LAN. Multiple first hop routers on the LAN combine to offer a single virtual first hop IPv6 router while sharing the IPv6 packet forwarding load [22].

The advantage of GLBP is that it additionally provides load balancing over multiple routers (gateways) using a single virtual IPv6 address and multiple virtual MAC addresses. The forwarding load is shared among all routers in a GLBP group rather than being handled by a single router while the other routers stand idle. Each host is configured with the same virtual IPv6 address, and all routers in the virtual router group participate in forwarding packets [12].

1. GLBP Addressing

A GLBP group allows up to four virtual MAC addresses per group. The AVG is responsible for assigning the virtual MAC addresses to each member of the group. Other group members request a virtual MAC address after they discover the AVG

through hello messages. Gateways are assigned the next MAC address in sequence. A virtual forwarder that is assigned a virtual MAC address by the AVG is known as a primary virtual forwarder. Other members of the GLBP group learn the virtual MAC addresses from hello messages. A virtual forwarder that has learned the virtual MAC address is referred to as a secondary virtual forwarder [22].

2. *GLBP Timers*

The default timers in GLBP are similar to HSRP and tunable too. The hello timer: Hello time default value 3 seconds. While the Hold time default value 10 seconds. These timers are tunable and are tuned to obtain minimum convergence and therefore making a network highly available [13].

V. SIMULATION AND RESULTS

This paper focuses on implementing first hop redundancy protocols in a network to increase the availability and reduce network downtime. The main objective is to implement different First hop redundancy protocols on three sites and compare the performance of each one. Each site connects to two different ISP to provide high availability, and if one of the links fails (connected to main ISP) the other will take over, this will provide a way to minimize network downtime, which is one of the most important goals of corporations in today's network.

A. Simulation Tool

GNS3 is a cross-platform graphical network simulator that runs on Windows, OS X, and Linux, it allows the combination of virtual and real devices, and is used to simulate complex networks without having dedicated network hardware such as routers and switches [14].

In GNS3 VPCS can provide Traffic Flow by using ICMP, TCP or UDP data flow. Simulating network fails IP Service Level Agreements (SLA) will be using and track object to help with failover process. IP SLAs is a network performance measurement and diagnostics tool that uses active monitoring. It is beneficial in the sense that it checks the reachability of a specific IP address and reports it back. Finally, results obtained are analyzed using Wireshark.

B. Network Design

The design used is a hierarchical design where each enterprise has two core layer routers and two access layer switches with partial mesh network topology in order to eliminate single points of failure in the enterprise network.

The design as shown in fig (2) consists of three enterprises (Enterprise 1, Enterprise 2, and Enterprise 3) each of them is connected to two ISP to disrepute internet access to the enterprises, each enterprise consists of two routers inside that connect internal network to the internet and two switches that provide layer 2 connectivity.
In order for the network to work and provide connectivity between the network nodes with fast convergence time OSPFv6 routing protocol is used to forward packets between the ISPs and the enterprises.

Each router has track object that is used to verify connection to ISP in case the connection goes down the track object decrements a value to the priority of active/master router which will make it have less priority than the standby/backup and it will result in making the standby/backup to become active/master router.

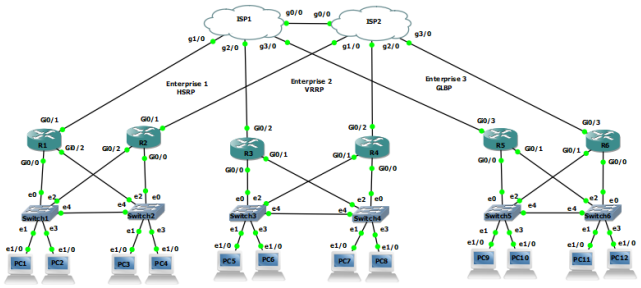


Fig. 2. Network Topology

C. Configuration

HSRPv6 Hot standby router protocol is configured on the first enterprise that contains R1 and R2. The link-local address is generated by entering standby IPv6 command. A link-local address is an IPv6 unicast address that can be automatically configured on any interface using the link-local prefix FE80: :/10, and the interface identifier in the modified EUI-64 format. VRRP Virtual Router Redundancy Protocol is configured on the second enterprise that contains R3 and R4. GLBP for IPv6 Gateway Load Balancing Protocol is configured for the third enterprise that contains R5 and R6.

IP SLA is Configured on enterprise routers to check the reachability of the ISP. If the reachability goes down it will report it back to the FHRP on the router using track object and bind it to the IP SLA when the ISP goes down. The track object will decrement a value of the priority of the router making it do to standby/backup while the other router becomes the active/master.

VI. RESULTS

This section will present and discuss the measurements taken in order to measure the performance of FHRPv6 and provide and analyze the results of each FHRPv6 then compare them.

- ***CPU Utilization***

CPU utilization is the sum of work handled by a Central Processing Unit. It is also used to estimate system performance. CPU utilization can vary according to the type and amount of computing task because some tasks require heavy CPU time while others require less CPU time.

CPU utilization shows the burden on a processor in terms of percentage that indicates if any changes are to be made in the system otherwise it may get exhausted of capacity.

- ***Traffic Flow Monitoring***

Traffic flow is a sequence of packets from a source to a destination, which may be another host, a multicast group, or a broadcast domain.

In GNS3 VPCS can provide Traffic Flow by using ICMP, TCP or UDP data flow.
This can be shown and analyzed by using protocol analysis programs like Wireshark.

Traffic Monitoring allows us to better plan and optimize the network for peak performance, when we configuring the network to support the correct amount of bandwidth or providing load balancing over multiple links, connectivity and speed are improved and users have a better experience when using the network.

Because of the different default hello and hold timers mentioned, this will cause inaccurate results, so two tests will be taken for FHRP one with the default timers and the other after the optimization of those timers.

A. HSRPv6

a) CPU Utilization without optimization

HSRPv6 consumed an average of 0.16% of CPU usage for R2 and 0.17% for R1 while it consumed an average of 0.16% of CPU usage for both routers.
HSRPv6 consumed an average of 24.5% of CPU utilization for both routers as shown in Fig (3).

R2

```
R2#sh processes cpu sorted
CPU utilization for five seconds: 25%/0%; one minute: 12%; five minutes: 14%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
  48        2308         114      20245  8.79%  0.73%  0.31%   0 Exec
   5        5703         306      18637  4.07%  0.84%  0.69%   0 Check heaps
 372       15370        3333       4611  2.39%  2.19%  2.02%   0 OSPFv3R-1/6/0
  81        1225          40      30625  1.83%  0.23%  0.13%   0 Per-minute Jobs
   3        7326        4430       1653  0.87%  0.86%  0.85%   0 OSPFv3H-1/6/0
 369        4199        4044       1038  0.87%  0.53%  0.51%   0 IPv6 Input
 364        4268       38951        109  0.79%  0.68%  0.57%   0 HSRP Common
 104        5976       10281        581  0.79%  0.82%  0.77%   0 VRRS Main thread
 269        5946       10277        578  0.71%  0.88%  0.78%   0 MMA DB TIMER
 281        5909       10284        574  0.55%  0.83%  0.77%   0 MMA DP TIMER
 370        1024         304       3368  0.47%  0.10%  0.09%   0 IPv6 ND
  84        4613        6706        687  0.47%  0.71%  0.65%   0 IOSv e1000
  96        2421        6706        361  0.47%  0.36%  0.33%   0 IOSv in console
  82        4601         747       6159  0.39%  0.49%  0.48%   0 Per-Second Jobs
 133        1465       19635         74  0.31%  0.20%  0.18%   0 IPAM Manager
 365        1282        1072       1195  0.23%  0.14%  0.15%   0 HSRP IPv6
  54         845         343       2463  0.15%  0.09%  0.08%   0 Net Background
 161         265        2887         91  0.07%  0.02%  0.00%   0 SSS Feature Time
 371        1965        2948        666  0.07%  0.05%  0.08%   0 MLD
```

R1

```
R1#sh processes cpu sorted
CPU utilization for five seconds: 41%/0%; one minute: 40%; five minutes: 35%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 373       107824      119174        904 20.23% 17.91% 15.64%   0 IP SLAs XOS Even
  48        11558         511      22618  9.59%  7.36%  2.67%   0 Exec
 372        15023        3348       4487  1.91%  1.98%  1.91%   0 OSPFv3R-1/6/0
 104         7697        9116        844  1.43%  1.33%  1.15%   0 VRRS Main thread
 281         7720        9115        846  1.19%  1.23%  1.11%   0 MMA DP TIMER
 269         7761        9117        851  1.11%  1.26%  1.10%   0 MMA DB TIMER
 364         8094       33780        239  0.79%  1.27%  1.15%   0 HSRP Common
  84         4754        5962        797  0.71%  0.77%  0.69%   0 IOSv e1000
   3         5714        4103       1392  0.55%  0.77%  0.73%   0 OSPFv3H-1/6/0
  96         3422        5961        574  0.47%  1.28%  0.64%   0 IOSv in console
 133         3009       17054        176  0.47%  0.51%  0.45%   0 IPAM Manager
 369         4088        3639       1123  0.39%  0.50%  0.52%   0 IPv6 Input
 371         2460        2673        920  0.31%  0.47%  0.45%   0 MLD
  82         3775         679       5559  0.31%  0.39%  0.39%   0 Per-Second Jobs
 161          542        2615        207  0.23%  0.16%  0.16%   0 SSS Feature Time
  27          649         708        916  0.23%  0.09%  0.08%   0 ARP Background
 365          961         987        973  0.23%  0.12%  0.13%   0 HSRP IPv6
 211          622         754        824  0.15%  0.09%  0.08%   0 ADJ background
  54          687         306       2245  0.15%  0.09%  0.08%   0 Net Background
--More--
```

Fig. 3. HSRPv6 CPU Utilization without Optimization

b) Traffic flow without Optimization

Fig (4) shows data traffic flow in HSRPv6 network throw R1. We can see from the following figure that the maximum bandwidth of the packet reached is 33 packets/sec.

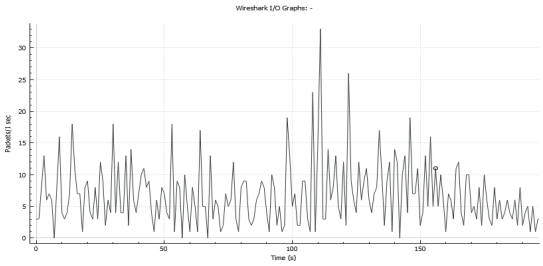


Fig. 4. HSRPv6 Traffic Flow without Optimization

c) CPU Utilization with Optimization

HSRPv6 consumed an average of 0.35 of CPU usage for R2 and 0.22% for R1 while it consumed an average of 0.28% of CPU usages for both routers.
HSRPv6 consumed an average of 29% of CPU utilization for both routers as shown in Fig (5).

R2

```
R2#sh processes cpu sorted
CPU utilization for five seconds: 53%/0%; one minute: 34%; five minutes: 25%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 372       23418        3654       6408 27.93% 13.08%  4.88%   0 OSPFv3R-1/6/0
   3        7665        6011       1275  6.94%  3.81%  1.62%   0 OSPFv3H-1/6/0
 369        5781        4591       1259  5.82%  2.66%  1.20%   0 IPv6 Input
 364        4889       21589        226  2.07%  1.56%  0.96%   0 HSRP Common
 269        4243        5853        724  1.83%  1.14%  0.79%   0 MMA DB TIMER
  48        4294         259      16579  1.67%  0.31%  0.58%   0 Exec
 281        4059        5851        693  1.35%  1.09%  0.74%   0 MMA DP TIMER
 104        4127        5849        705  1.19%  0.94%  0.71%   0 VRRS Main thread
  82        3308         472       7008  0.79%  0.67%  0.55%   0 Per-Second Jobs
 133        1244       10881        114  0.63%  0.37%  0.23%   0 IPAM Manager
  96        1559        3915        398  0.55%  0.35%  0.28%   0 IOSv in console
 371        1631        1830        891  0.47%  0.46%  0.35%   0 MLD
 365        1484        1347       1101  0.39%  0.38%  0.28%   0 HSRP IPv6
  84        2362        3915        603  0.23%  0.48%  0.42%   0 IOSv e1000
--More--
```

R1

```
R1#sh processes cpu sorted
CPU utilization for five seconds: 36%/0%; one minute: 37%; five minutes: 33%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 373       53870       59459        906 18.96% 18.60% 11.62%   0 IP SLAs XOS Even
 364        6084       17483        347  2.24%  2.16%  1.28%   0 HSRP Common
 372        8523        1607       5303  1.83%  2.10%  1.60%   0 OSPFv3R-1/6/0
 104        4377        4800        911  1.68%  1.45%  0.93%   0 VRRS Main thread
  57        7031         365      19263  1.52%  1.45%  1.15%   0 TTY Background
  48        3208         247      12987  1.44%  0.75%  0.58%   0 Exec
 269        4324        4801        900  1.20%  1.36%  0.91%   0 MMA DB TIMER
 281        4429        4804        921  1.12%  1.44%  0.90%   0 MMA DP TIMER
 369        3269        2376       1375  0.95%  1.02%  0.67%   0 IPv6 Input
   3        3779        2041       1851  0.79%  0.96%  0.71%   0 OSPFv3H-1/6/0
  84        2393        3227        741  0.79%  0.72%  0.48%   0 IOSv e1000
 133        1728        8821        195  0.47%  0.58%  0.39%   0 IPAM Manager
 371        1610        1537       1047  0.39%  0.25%  0.21%   0 MLD
  82        2537         397       6390  0.39%  0.43%  0.41%   0 Per-Second Jobs
  96        1339        3227        414  0.23%  0.38%  0.27%   0 IOSv in console
 278         492         387       1271  0.23%  0.10%  0.08%   0 Crypto IKEv2
 120        1227          95       1291  0.23%  0.06%  0.11%   0 CDP Protocol
 370         654         165       3963  0.15%  0.09%  0.09%   0 IPv6 ND
 195         367         370        991  0.15%  0.11%  0.08%   0 RUDPV1 Main Proc
 247         283         707        400  0.15%  0.07%  0.04%   0 CCE DP URLF cach
 365        1018        1025        993  0.15%  0.31%  0.20%   0 HSRP IPv6
 102         357         375        952  0.15%  0.10%  0.08%   0 FHRP Main thread
  56         168         134       1253  0.15%  0.02%  0.00%   0 Logger
  54         520         147       3537  0.15%  0.13%  0.08%   0 Net Background
--More--
```

Fig. 5. HSRPv6 CPU Utilization with Optimization

d) Traffic flow with Optimization

Fig (6) shows data traffic flow in HSRP network through R1. We notice in the following figure that the maximum bandwidth of the packet still 33 packets/sec.

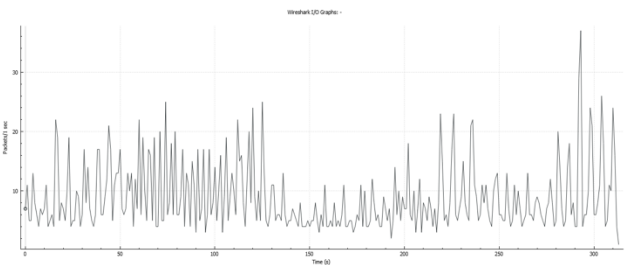


Fig. 6. HSRPv6 Traffic flow with Optimization

B. VRRPv6

a) CPU Utilization without optimization

VRRP consumed an average of 0.85% of CPU usage for R4 and 0.55% for R3 while it consumed an average of 0.70% of CPU usage for both routers.
VRRP consumed an average of 23.5% of CPU utilization for both routers as shown in Fig (7).

R4

```
R4#sh processes cpu sorted
CPU utilization for five seconds: 8%/0%; one minute: 11%; five minutes: 11%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 370       72530       16838       4307  1.19%  1.40%  1.50%   0 OSPFv3R-1/6/0
 103       29319       14651       2001  0.87%  0.94%  0.74%   0 VRRP Main thread
 281       52684       71997        731  0.79%  1.04%  1.09%   0 MMA DP TIMER
 367       29745       25831       1151  0.79%  0.57%  0.60%   0 IPv6 Input
  84       37010       45920        805  0.71%  0.75%  0.77%   0 IOSv e1000
 104       53096       72013        737  0.63%  0.95%  1.07%   0 VRRS Main thread
  82       28373        4734       5993  0.55%  0.49%  0.51%   0 Per-Second Jobs
  48        4934         591       8348  0.47%  0.60%  0.33%   0 Exec
  96       18541       45920        403  0.39%  0.35%  0.37%   0 IOSv in console
 269       52866       72002        734  0.39%  1.32%  1.30%   0 MMA DB TIMER
   3       29611       20094       1473  0.23%  0.48%  0.52%   0 OSPFv3H-1/6/0
  54        5316        2725       1950  0.23%  0.09%  0.11%   0 Net Background
 369       15333       18879        812  0.23%  0.14%  0.14%   0 MLD
  27        4356        4922        885  0.23%  0.07%  0.06%   0 ARP Background
 368        6105        1992       3064  0.15%  0.12%  0.14%   0 IPv6 ND
 102        3975        4687        848  0.15%  0.11%  0.10%   0 FHRP Main thread
 211        4697        5448        862  0.15%  0.10%  0.09%   0 ADJ background
--More--
```

R3

```
R3#sh processes cpu sorted
CPU utilization for five seconds: 34%/0%; one minute: 36%; five minutes: 36%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 371      940648      932696       1008 18.87% 21.20% 21.63%   0 IP SLAs XOS Even
  81        5979         166      36018  2.07%  0.25%  0.14%   0 Per-minute Jobs
 269       63425       67472        940  1.75%  1.46%  1.41%   0 MMA DB TIMER
 104       63844       67484        946  1.75%  1.39%  1.41%   0 VRRS Main thread
   5       22300        1296      17206  1.59%  0.63%  0.48%   0 Check heaps
 281       63121       67452        935  1.35%  1.43%  1.47%   0 MMA DP TIMER
  48        6193         867       7143  0.95%  2.16%  1.16%   0 Exec
  84       41474       43038        963  0.79%  0.97%  0.95%   0 IOSv e1000
 133       27239      130507        208  0.71%  0.68%  0.65%   0 IPAM Manager
  82       23481        4469       5254  0.55%  0.51%  0.55%   0 Per-Second Jobs
 370       71384       13940       5120  0.55%  1.19%  1.28%   0 OSPFv3R-1/6/0
   3       33547       17103       1961  0.47%  0.71%  0.69%   0 OSPFv3H-1/6/0
 367       33744       24429       1381  0.39%  0.67%  0.69%   0 IPv6 Input
 103       38853       15356       2530  0.31%  0.51%  0.84%   0 VRRP Main thread
 120        3575         941       3799  0.23%  0.04%  0.05%   0 CDP Protocol
 195        4027        4412        912  0.15%  0.09%  0.08%   0 RUDPV1 Main Proc
--More--
```

Fig. 7. VRRP CPU Utilization without Optimization

b) Traffic flow without Optimization

Fig (8) shows data traffic flow in VRRP network throw R3. We note in the following figure that the maximum bandwidth of the packets is 15.5 packets/sec.

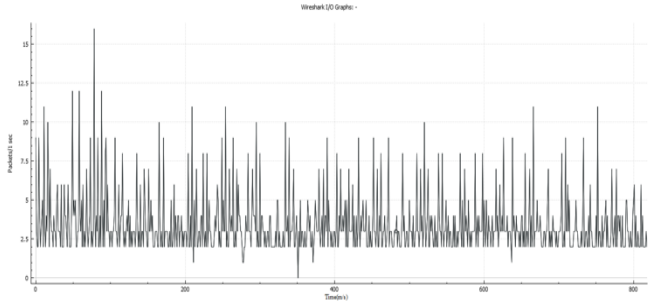


Fig. 8. VRRP Traffic flow without Optimization

c) CPU Utilization with Optimization

VRRP consumed an average of 1.70% of CPU usage for R4 and 0.93% for R3 while it consumed an average of 1.31% of CPU usage for both routers. VRRP consumed an average of 27% of CPU utilization for both routers as shown in Fig (9).

R4

```
R4#sh processes cpu sorted
CPU utilization for five seconds: 26%/0%; one minute: 17%; five minutes: 20%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
  48        5033         364      13826 14.16%  3.64%  1.21%   0 Exec
 103        5318        2268       2344  2.00%  1.98%  1.13%   0 VRRP Main thread
 281        4347        5439        799  1.91%  1.27%  0.83%   0 MMA DP TIMER
 370        9787        1679       5829  1.75%  1.45%  1.50%   0 OSPFv3R-1/6/0
 367        4386        2927       1498  1.04%  0.94%  0.78%   0 IPv6 Input
  84        2382        3649        652  0.79%  0.72%  0.51%   0 IOSv e1000
 269        4371        5435        804  0.79%  0.85%  0.65%   0 MMA DB TIMER
 104        4405        5440        809  0.71%  1.12%  0.79%   0 VRRS Main thread
  96        1472        3649        403  0.71%  0.39%  0.30%   0 IOSv in console
 369        1610        1709        942  0.47%  0.48%  0.39%   0 MLD
   3        3971        2027       1959  0.47%  0.65%  0.63%   0 OSPFv3H-1/6/0
--More--
```

R3

```
R3#sh processes cpu sorted
CPU utilization for five seconds: 34%/0%; one minute: 38%; five minutes: 34%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 371       40841       39228       1041 20.07% 19.12%  9.53%   0 IP SLAs XOS Even
  48        3960         389      10179  2.55%  1.37%  0.81%   0 Exec
 269        3158        3332        947  1.91%  1.34%  0.71%   0 MMA DB TIMER
 281        3218        3335        964  1.67%  1.41%  0.76%   0 MMA DP TIMER
 104        3295        3337        987  1.19%  1.53%  0.80%   0 VRRS Main thread
 367        2875        2122       1354  1.03%  0.94%  0.61%   0 IPv6 Input
 370        6968        1233       5651  1.03%  1.41%  1.32%   0 OSPFv3R-1/6/0
  84        1521        2307        659  0.63%  0.78%  0.38%   0 IOSv e1000
 369        1297        1187       1092  0.47%  0.45%  0.31%   0 MLD
 133        1241        6058        204  0.39%  0.63%  0.33%   0 IPAM Manager
   3        2827        1609       1756  0.39%  0.69%  0.58%   0 OSPFv3H-1/6/0
  82        2107         307       6863  0.39%  0.48%  0.38%   0 Per-Second Jobs
 195         257         283        908  0.23%  0.11%  0.06%   0 RUDPV1 Main Proc
 103        4976        1728       2879  0.23%  1.49%  1.07%   0 VRRP Main thread
  99         329         290       1134  0.23%  0.13%  0.07%   0 mDNS
 368         649         147       4414  0.15%  0.13%  0.10%   0 IPv6 ND
--More--
```

Fig. 9. VRRP CPU Utilization with Optimization

d) Traffic flow with Optimization

Fig (10) shows data traffic flow in VRRP network throw R3. We can see in the following diagram that the maximum bandwidth of the packets is 25 packets/sec.

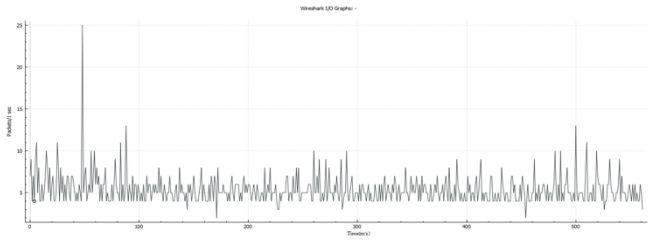


Fig. 10. VRRP Traffic flow with Optimization

C. GLBPv6

a) CPU Utilization without optimization

GLBP consumed an average of 0.60% of CPU usage for R6 and 0.88% for R5 while it consumed an average of 0.74% of CPU usage for both routers.
GLBP consumed an average of 25.5% of CPU utilization for both routers as shown in Fig (11).

R6

```
R6#sh processes cpu sorted
CPU utilization for five seconds: 19%/0%; one minute: 17%; five minutes: 15%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
   3       12883         966      13336  6.71%  3.45%  1.51%   0 Exec
 372       20449        4245       4817  2.47%  2.77%  2.23%   0 OSPFv3R-1/6/0
 104       10648       14732        722  1.59%  1.04%  1.02%   0 VRRS Main thread
 375        9952        5719       1740  1.27%  1.07%  0.97%   0 OSPFv3H-1/6/0
 281       10579       14732        718  1.19%  1.13%  1.02%   0 MMA DP TIMER
 269       10719       14730        727  1.11%  1.09%  1.01%   0 MMA DB TIMER
 370        4404       28619        153  0.87%  0.51%  0.42%   0 GLBP
  82        6710        1041       6445  0.79%  0.77%  0.77%   0 Per-Second Jobs
 368        6285        5191       1210  0.79%  0.77%  0.65%   0 IPv6 Input
  84        7654        9533        802  0.71%  0.80%  0.78%   0 IOSv e1000
  54        1182         586       2017  0.23%  0.11%  0.08%   0 Net Background
 133        2096       28243         74  0.15%  0.23%  0.21%   0 IPAM Manager
  96        4607        9533        483  0.15%  0.37%  0.39%   0 IOSv in console
--More--
```

R5

```
R5#sh processes cpu sorted
CPU utilization for five seconds: 39%/0%; one minute: 37%; five minutes: 36%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 373      183444      182190       1006 22.95% 21.34% 20.09%   0 IP SLAs XOS Even
 372       18012        4087       4407  3.59%  2.35%  2.07%   0 OSPFv3R-1/6/0
   3        8678         795      10915  1.27%  0.88%  0.91%   0 Exec
 104       12987       13637        952  1.27%  1.41%  1.37%   0 VRRS Main thread
 269       12971       13636        951  1.19%  1.46%  1.40%   0 MMA DB TIMER
 281       12905       13642        945  1.19%  1.50%  1.39%   0 MMA DP TIMER
  84        8945        8841       1011  1.11%  1.12%  1.01%   0 IOSv e1000
 376        7780        5206       1494  0.95%  0.95%  0.87%   0 OSPFv3H-1/6/0
 368        5038        4361       1155  0.87%  0.53%  0.50%   0 IPv6 Input
 370        8325       26488        314  0.79%  0.93%  0.92%   0 GLBP
  82        5283         970       5446  0.63%  0.65%  0.59%   0 Per-Second Jobs
--More--
```

Fig. 11. GLBP CPU Utilization without Optimization

b) Traffic flow without Optimization

Fig (12) shows data traffic flow in GLBP network throw R5 without Optimization.

We note in the following figure that the maximum bandwidth of the packets in R5 is 16.5 packets/sec.

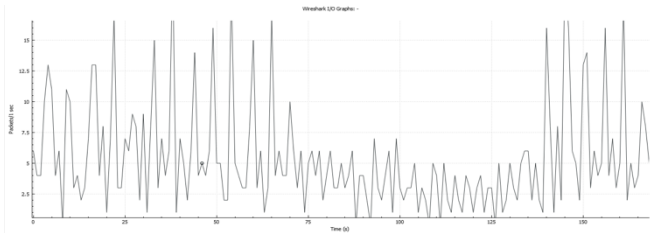
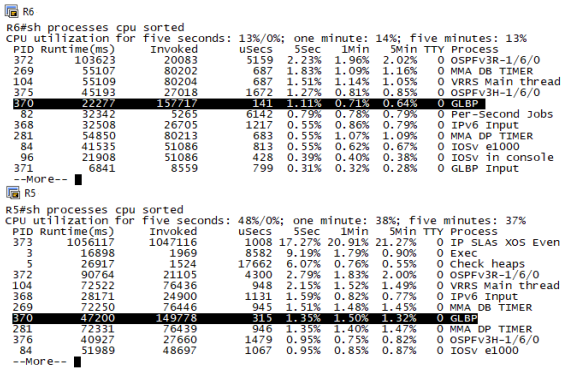


Fig. 12. GLBP Flow traffic without Optimization

c) CPU Utilization with Optimization

GLBP consumed an average of 0.82% of CPU usage for R6 and 1.39% for R5 while it consumed an average of 1.10% of CPU usage for both routers.

GLBP consumed an average of 25% of CPU utilization for both routers as shown in Fig (13).



```
R6
R6#sh processes cpu sorted
CPU utilization for five seconds: 13%/0%; one minute: 14%; five minutes: 13%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 372      103623       20083       5159  2.23%  1.96%  2.02%   0 OSPFv3R-1/6/0
 269       55107       80202        687  1.83%  1.09%  1.16%   0 MMA DB TIMER
 104       55109       80204        687  1.51%  1.14%  1.05%   0 VRRS Main thread
 375       45193       27018       1672  1.27%  0.81%  0.85%   0 OSPFv3H-1/6/0
 370      222277      152217        141  1.11%  0.41%  0.64%   0 GLBP
  82       32342        5265       6142  0.79%  0.78%  0.79%   0 Per-Second Jobs
 368       32508       26705       1217  0.55%  0.86%  0.79%   0 IPv6 Input
 281       54850       80213        683  0.55%  1.07%  1.09%   0 MMA DP TIMER
  84       41535       51086        813  0.55%  0.62%  0.67%   0 IOSv e1000
  96       21908       51086        428  0.39%  0.40%  0.38%   0 IOSv in console
 371        6841        8559        799  0.31%  0.32%  0.28%   0 GLBP Input
  --More--
R5
R5#sh processes cpu sorted
CPU utilization for five seconds: 48%/0%; one minute: 38%; five minutes: 37%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 373      1056117     1047116       1008 17.27% 20.91% 21.27%   0 IP SLAs XOS Even
   3        16898        1969       8582  9.19%  1.79%  0.90%   0 Exec
   5        26917        1524      17662  6.07%  0.76%  0.55%   0 Check heaps
 372        90764       21105       4300  2.79%  1.83%  2.00%   0 OSPFv3R-1/6/0
 104        72522       76436        948  2.15%  1.52%  1.49%   0 VRRS Main thread
 368        28171       24900       1131  1.59%  0.82%  0.77%   0 IPv6 Input
 269        72250       76446        945  1.51%  1.48%  1.45%   0 MMA DB TIMER
 370        47200      149778        315  1.35%  1.50%  1.32%   0 GLBP
 281        72331       76439        946  1.35%  1.40%  1.47%   0 MMA DP TIMER
 376        40927       27660       1479  0.95%  0.75%  0.82%   0 OSPFv3H-1/6/0
  84        51989       48697       1067  0.95%  0.85%  0.87%   0 IOSv e1000
  --More--
```

Fig. 13. GLBP CPU Utilization with Optimization

d) Traffic flow with Optimization

Fig (14) shows data traffic flow in GLBP network throw R5

We can see in the figure below that the maximum bandwidth of the packets in R5 is 44 packets/sec.

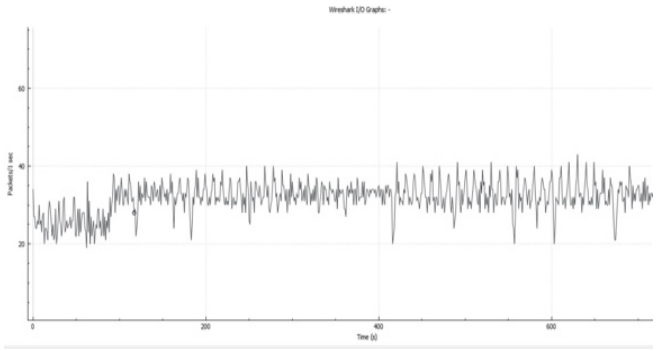


Fig. 14. GLBP Traffic Flow with Optimization

VII. COMPARISON

A. CPU Utilization

Fig (15) show CPU Utilization comparison between FHRP.

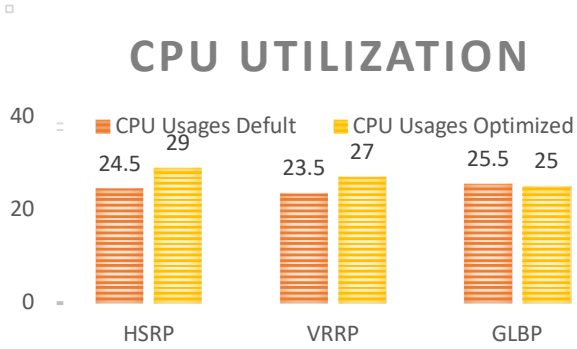


Fig. 15. FHRP CPU Utilization Comparison Comparison

We can see from the figure that VRRP does not require high CPU consumption, because it sends the packets advertisement interval time every 1000m/s, thus we conclude that VRRP has the best Utilization of CPU before Optimization. While GLBP has the best utilization of CPU after optimization, because GLBP works on the principle of load balancing between routers, unlike HSRP and VRRP.

B. Packets Loss

Fig (16) show Packet loss comparison between FHRP during convergence time. GLBP has the lowest packet loss after optimization.

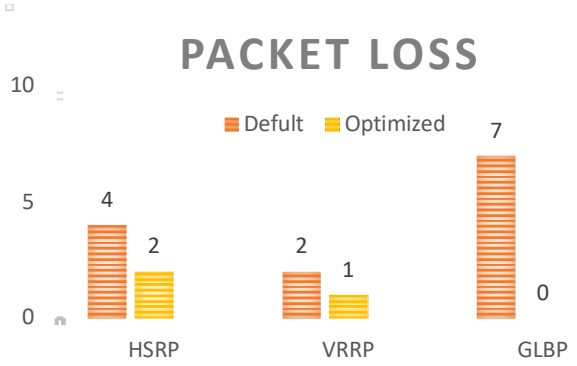


Fig. 16. FHRP Packet Loss Comparison

VIII. CONCLUSION

After implementing, optimizing and testing the different FHRP whilst Studying and analyzing their output of the conversion time, CPU utilization, and traffic flow it is clear to see that GLBP has higher performance than HSRP and VRRP. all the load balancing futures make GLBP an efficient and reliable protocol or redundancy and providing more availability to the network, but the only downside is that GLBP is cisco proprietary so it only runs on cisco devices.

Therefore, it can be concluded the GLBP is better than HSRP and VRRP in terms of performance and achieving higher availability in the network.

REFERENCES

[1] Chris Oggerino, "High Availability Network Fundamentals", Cisco Press, 1st edition, 2001.

[2] Priscilla Oppenheimer, "Top-Down Network Design", Cisco Press, 3ed edition, 2010.

[3] Richard Froom, "Implementing Cisco IP Switched Networks (SWITCH) Foundation Learning Guide: Foundation Learning for SWITCH 642- 813", Cisco Press, 1st edition, 2010.

[4] Rolf McClellan, Nick Lippis "Network-Level Redundancy/Resilience for High-Availability Campus LANs"," ZDTag white paper", 1999, pp 6-8.

[5] Mike Miclot, John Mower, "Reducing the Risk, Cost and Frequency of Production Stoppages Using Network Redundancy", 2010.

[6] Priyanka Dubey, Shilpi Sharma, Aabha Sachdev, "Review of First Hop Redundancy Protocol and Their Functionalities", International Journal of Engineering Trends and Technology, 2013, pp 1085- 1088.

[7] T. Li, B. Cole, P. Morton, D. Li," Cisco Hot Standby Router Protocol", Request for Comments: 2281, 1998.

[8] cisco.com: Hot Standby Router Protocol Features and Functionality, http://www.cisco.com/c/en/us/support/docs/ip/hot-standby-router-protocol-hsrp/9234-hsrpguidetoc.html, 2006.

[9] S. Knight, D. Weaver, D. Whipple, R. Hinden, D. Mitzel, P. Hunt, P. Higginson, M. Shand, A. Lindem, "Virtual Router Redundancy Protocol", Request for Comments: 2338,1998.

[10] R. Hinden, Ed, "Virtual Router Redundancy Protocol", Request for Comments: 3768, 2004.

[11] S. Nadas, "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", Request for Comments: 5798, 2010.

[12] cisco.com: GLBP - Gateway Load Balancing Protocol , http://www.cisco.com/en/US/docs/ios/12_2t/12_2t15/feature/guide/ft_glbp.html, 2016.

[13] First Hop Redundancy Protocol (FHRP)". Cisco. Retrieved 2019-06-05

[14] Jason C. Neumann, "The Book of GNS3", No Starch Press, 1st edition, 2015

[15] A. Zemtsov "Performance Evaluation of First Hop Redundancy Protocols for a Computer Networks of an Industrial Enterprise" 2019 International Multi-Conference on Industrial Engineering and Modern Technologies. 1-4 Oct. 2019.

[16] M. Mansour, A.Ghneimat, R.Alasem, F. Jarray, "Performance Analysis and Functionality Comparison of First Hop Redundancy Protocol", Journal of Ubiquitous Systems & Pervasive Networks. Volume 15, No. 1 (2021) pp. 49-58

[17] M. Mansour "Performance Evaluation of First Hop Redundancy Protocols ", The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) November 2-5, 2020, Madeira, Portugal.

[18] Usman Anwar , Jing Teng ; Hafiz Ahmad Umair ; Ammar Sikander "Performance Analysis and Functionality Comparison of FHRP Protocols "IEEE 11th International Conference on Communication Software and Networks (ICCSN), 12-15 June 2019.

[19] Z. U. Rahman et al., "Performance Evaluation of First Hop Redundancy Protocols," J. Appl. Environ. Biol. Sci., vol. 7, no. 3, pp. 268–278, 2017.

[20] Imelda Ristanti Julia et al, "Protocol (FHRP) on VRRP, HSRP, GLBP with Routing Protocol BGP and EIGRP", The 8th International Conference on Cyber and IT Service Management (CITSM 2020) On Virtual, October 23-24, 2020.

[21] [Andy Sholomon, Tom Kunath, "Enterprise Network Testing", Cisco Press, 1st edition, 2011.

[22] The cost of network Downtime (And how to prevent it) – guardian computer https://www.gcit.net/blog/cost-of-network-downtime.

[23] Configuring First Hop Redundancy Protocols in IPv6 https://www.cisco.com/en/US/docs/iosxml/ios/ipv6/configuration/xe-3sg/ip6-fhrp.html#GUID-DCB20ADF-1F8E-434B-AE97-54802879F34F.

[24] VRRPV3,. http://docs.ruckuswireless.com/fastiron/08.0.60/fastiron-08060-l3guide/GUID-D3E29B82-AE92-4661-8BF6EF7AA8E99139.html.

[25] Cisco.com: Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide, http://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/5_x/nx-os/unicast/configuration/guide/l3_cli_nxos/l3_vrrp.html, 2016.

The Internet of Things : Communication Technologies, Architecture and Future Applications of a smart connected world: A review

Hamza Abedulsalam Saoud[1*], Abubaker Salem Ahmed[1]
[1]*Libyan Center for Engineering Research and Information Technology, Bani -Walid, Libya.*
hamza.fughi@gmail.com[1*], Bker79sm@gmail.com

***Abstract*— Internet of things (IoT) is a new paradigm getting very popular day by day that has changed the traditional way of living into a high tech life style. Smart city, smart homes, energy saving, smart transportation. The very reason for this to happen is the advancement in technology and its ability to get linked to everything. The Internet of Things is driven by an expansion of the Internet with an ability to provide smarter services to the environment as more data becomes available. In the future, hundreds of billions of smart sensors and devices will interact with one another without human intervention, on a Machine-to-Machine (M2M) basis. They will generate an enormous amount of data at an unprecedented scale and resolution, providing humans with information and control of events and objects even in remote physical environments. This paper aims to provide a comprehensive overview of what the Internet of Things is, communication technologies, and the possible applications of IoT in the future, and a review on privacy and security. This paper will help readers and researchers understand the Internet of Things and its applicability in the real world.**

Keywords*— *Internet of things, Machine-to-Machine, IoT applications, communication technologies.

I. INTRODUCTION

With the rapid advancements in technology, the internet of things IoT is coming down the road which is burgeoning as an ubiquitous global computing network where everyone and everything will be connected to the Internet in different environments including homes, aerospace and various transportations[1]. Controlling systems and IoT combination is one of the main concerns of researchers. Different approach has been proposed to control IoT devices. IoT security has the highest priority concerns and became the first topic for research in the field of IoT technology [2]. The year 1999 was easily one of the most significant for the IoT history, as Kevin Ashton coined the concept "the internet of things." where he described IoT as a technology that connected several devices with the help of a member the Radio Frequency Identification (RFID) development community in 1999, and it has recently become more relevant to the practical world largely because of the growth of mobile devices, embedded and ubiquitous communication, cloud computing and data analytics [3]. The basic idea of IoT is to allow autonomous and secure connection and exchange of data between real world devices and applications, that will enable a networking infrastructure that connects a large number of devices to allow them to collect data and communicate with each other in order to make processed smart decisions [4]. The IoT sensors enable the physical objects to observe, hear, think and react to share data to make these decisions. The transformation from traditional objects to smart objects be done by the technologies of IoT like ubiquitous computing, wireless sensor networks, Internet protocols [5]. Meanwhile, a number of challenges are in the way of the IoT. In terms of scalability, IoT applications that require large numbers of devices are often difficult to implement because of the restrictions on time, memory, processing, and energy constraints [6]. Fig.1, shows that with the IoT technology, anything's we will able to communicate to the internet at any time from any place to provide any services by any network to anyone [7]. The IoT sensors can use various types of connections like Wi-Fi, and Bluetooth, in the addition to allowing large-scale area connectivity used many technologies such as 3G, 4G, 5G ,GSM, GPRS Networks.

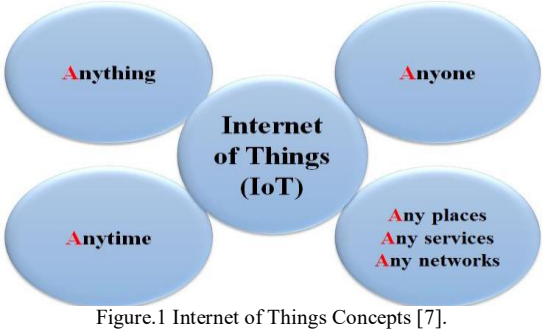


Figure.1 Internet of Things Concepts [7].

According to statistica [8] website, It is expected that the total number of smart devices connected together will reach up to 50 billion. According to Siemens research, up to 2020, near about 26 billion physical objects will be connected together on the internet. They can communicate each other and forwarding and process required data on the cloud [9]. Also, in 2025, the total number of connected devices in the world will be approximately 75.44 billion, Currently, there are about 50 billion devices interconnected by the IoT which would ragingly grow to about 80 billion by the year 2030. Fig. 2 shows the rapid growth of IoT by 2030.

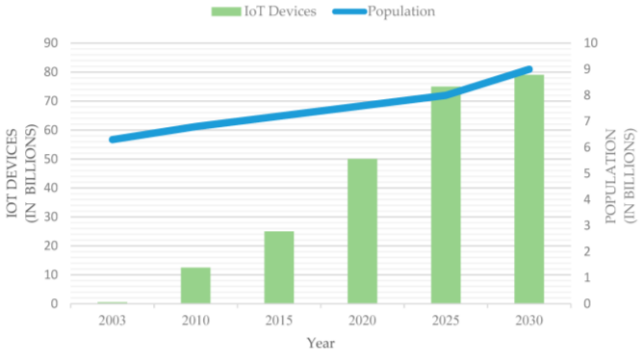


Figure 2. (IoT) connected devices from 2003 to 2030 (in billions).

The main factor of this growth is not the population of the world but the smart devices connected to the internet. The integrated technologies are playing big role to connect the physical things together and exchange the information among them [9]. This environment where the machine can talk to another machine (Machine-to-Machine) (M2M) learning and human can talk to machine [10].

The purpose of this paper is to present the IoT communication technology, its engineering architecture and

its possible future applications, a review the security and privacy of the (IOT) for connected devices.

II. Generic Architecture of Internet of Things

Implementation of IoT systems is based on an architecture consisting of several layers: from the field data acquisition layer at the bottom to the application layer at the top have the ability to perform interoperability among heterogeneous assets around us[11]. IoT system architecture should be adaptive to make devices interact with other dynamically and automatically that support connection with each other [12]. A brief introduction to the layers of IoT system architecture is as follows;

A. Sensing layer

This is the lowest layer of IoT architecture. this layer consists of sensor networks, embedded systems and readers or other soft sensors in different forms. The sensors enable the interconnection of the physical and digital worlds allowing real-time information to be collected and processed In the sensing layer, the smart systems on tags or sensors are able to automatically sense the environment and exchange data between these devices. Things can be uniquely identified and the surrounding environments can be monitored for various purposes and applications. Other functions for this layer, identification and information storage (e.g. RFID tags), information collection (e.g. wireless sensor networks), information processing (e.g. embedded edge processors), communication, control and actuation [13] .

B. Gateways and Network layer

This layer (also known as transmission layer) act as a bridge to transfer incoming digitized data from object sensing layer to the service management layer through the transmission mediums like WiFi, Bluetooth, WiMaX, Zigbee, GSM, GPRS, 3G/4G technologies with protocols like IPv4, IPv6[11]. Current networks, often tied with very different protocols, have been used to support machine-to-machine (M2M) networks and their applications [14]. The network layer in IoT, connects all things and allows them to be aware of their surroundings. Via the network layer, things can share data with the connected things, which is crucial to intelligent event management and processing in IoT. The layer performs other functions; Gateway - Routing & data addressing - Security and privacy- Error detection and Correction [15].

C. Service Management layer

The service layer is important for the Internet of Things. It is a middleware layer that provides a service to its requesters after specifying the names and address and is responsible for managing the service and has a link with the important database [16]. IoT devices implement different types of services, where each device only communicates with those other devices that perform the same type of service. performs the following functions; Service storage & orchestration – Service composition & organization – Virtual Entity resolution – IoT service –IoT service resolution – VE & IoT service monitoring, EPC (Electronic Product Code) information service and ONS (Object Naming Service) [17].

D. Interface layer

Includes the IoT application. This layer is at the top of the architecture and is responsible for delivery of various applications to different users in system IoT [17]. The applications of IoT can be covers “smart” environments/ spaces in domains such as smart homes, smart cities, smart health, animal tracking, etc. food and drug etc. With the increasing maturity of RFID technology, numerous applications are evolving which will be under the umbrella of IoT. It has the responsibility to provide the services to the applications [18]. The services may be varying for each application because services depend on the information that is collected by sensors to simplify the management and interconnection of things. Fig.2 below shows the four layers that system IoT architecture represents.

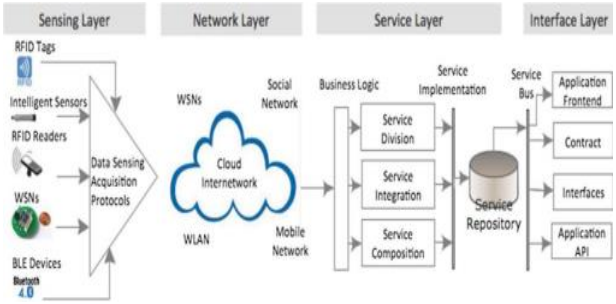


Figure 2. Architectural Layers of IoT [13].

III. Key Communication Technologies Involved in IoT

In IoT applications, it is mandatory to transmit the data generated by the devices or sources to the internet. They should be easy to locate, identify, address, and control [19]. In this section we discuss the technological developments that can help in the large-scale development of IoT:

A. RFID (radio frequency identification)

RFID is first technology that exists from the time when the concept of IoT technology was found in late 90‟s [20]. RFID systems mainly focuses on Near Field Communication (NFC) and Radio-Frequency identification system. RFID systems supports IoT networks by ensuring that IoT objects are equipped with identifiers and smart tags that make the objects manageable. Furthermore, RFID technology enables the IoT objects to have smart chips that give the objects the ability to sense information in their environment, compute and communicate with other Humans or beings. This wireless technology consists of two components: tags and readers. Tags, which use radio waves to communicate their identity and other information to nearby readers, can be passive or active. Passive RFID tags are powered by the reader and contain no batter. The transducer includes transmitter and responder, and we can identify tags in two forms [20].

B. Cloud Computing platforms

An IoT cloud is a type of Internet-based computing that supports IoT devices and applications. This includes the underlying infrastructure, servers and storage, needed for real-time operations and processing. Cloud can offer an effective solution for IoT service management and composition as well as for implementing applications and

services that exploit the things or the data produced by them. In the past, this whole process was conducted on a personal computer or local server. In the present, the cloud can benefit from IoT by extending its scope to deal with real-world things in a more distributed and dynamic[21]. Moreover, IoT cloud platform services provide more flexibility in case you want to limit storage requirements that produce large volume of data collected at different speeds, and a viable solution for managing data produced by IoT is the cloud [5].

Figure. 3. A typical Cloud Computing Scenario

C. *Wireless sensor networks (WSNs)*

Wireless sensor networks (WSNs) are other essential technologies for collecting data in IoT environment[16]. Wireless sensor networks refer to networks of spatially dispersed and dedicated sensors that monitor and record the physical conditions of the environment and forward the collected data to a central location. WSNs have many nodes and each node has four parts: sensors, battery, microcontroller and memory They are generally composed of low power radios, several smart sensors and embedded CPUs (Central Processing Units). These devices are used to form wireless sensor network (WSN) which is necessary to provide sensing services and to monitor environmental conditions. the major components that made the WSN monitoring network includes [5]:

a)- Hardware used for WSN - Typically contains the interfaces for sensors, processing units, transceivers and power supplying units.

b)- The communication protocol stack- Nodes in a WSN network need to communicate between the nodes to transmit the data in hop-by-hop basis or multi-hop basis to a base station [5]. There are types of WSNs mobile WSNs, multimedia WSNs, underground WSNs, or underwater WSNs[22].

D. *5G / 6G Wireless Communications Networks*

5G wireless networks provide the connectivity for IoT system. 5G will increase cellular bandwidth by huge amounts, making it much easier for the Internet of Things to network large numbers of devices together. Also, With 5G technology , data-transfer speeds will increase significantly, this increase in speed will allow IoT devices to communicate and share data faster than ever. There are lot of requirements of IoT applications such as high data rates, high scalability, low latency, reliability, security, and mobility, provided by the 5G network [23]. However, it opens a set of challenges on the architecture of 5G-IoT that need to be considered are: scalability, security assurance and privacy concerns. On the other hand, 6G wireless communication networks are envisioned to revolutionize customer services and applications via the Internet of Things (IoT) towards a future of fully intelligent and autonomous systems. 6G wireless networks will be able to use higher frequencies than 5G networks and provide substantially higher capacity and much lower latency. This is 1,000 times faster oared expected to empower future IoT networks, including Artificial Intelligence, Terahertz communications, massive ultra-reliable and low-latency communications, and blockchain systems and Industrial Internet of Things applications [24].

E. *Machine-to-Machine (M2M) communication*

IoT is also equally plays huge role in today's integration of machine and man like Machine-to-Machine (M2M). M2M, is exactly as it sounds: two machines "communicating" or exchanging data . This communication paradigm facilitates ubiquitous communications with full mechanical automation, where a large number of intelligent devices connected by wired/wireless links, interact with each other without direct human intervention [25]. The main components of an M2M networks include sensors, RFID, a Wi-Fi technology, Bluetooth, cellular communications, and autonomic computing software programmed to help a network device determine data and make decisions[6]. M2M communication technology is often used for remote monitoring. The detailed difference between IoT and M2M is depicted in Table 1.

TABLE 1. Difference between IoT and M2M

IoT	M2M
IoT connects a computer with "things", systems and people.	It connects "things" to a computer [26]
Horizontal connection IoT uses. objects to a cloud communication	It supports Point-to-point communication (vertical)
Hardware and Software based technology	M2M communication is primarily oriented towards an hardware [27]
Active internet connection is needed in most of the cases	Devices do not necessarily rely on an internet connection
Unlimited integration options	limited integration options

F. *6LoWPAN*

6LoWPAN (IPv6 over Low Power Wireless Personal Area Network) is the first and most common communication protocol for the IoT. 6LoWPAN system provides the upper layer system for use with low power wireless communications for IoT and M2M, originally intended for IEEE 802.15.4 protocol which cannot exceed 128 bytes, it is now used with many low power wireless mesh network where every node has its own IPv6 address[28]. This allows the node to connect directly with the Internet using open standards. Thus, it is able to provide a direct connection to a huge variety of networks including direct connection to the Internet. It also supports different length addresses, low bandwidth, different topologies including star or mesh, power consumption, low cost, scalable networks, mobility, unreliability [29].

G. Optical Wireless Technologies(WSNs)

Photonic technologies have played and will play in the future a significant role in the development of the IoT. in the other hand, optical technologies have the potential to greatly help in realizing future smart infrastructures and systems they provide fundamental components and high-performance communication and network capabilities[30]. Moreover, Rapid developments in the field of optical technologies in the form of technologies like Li-Fi (wireless optical networking technology) and optical bidirectional technology (BiDi) could be a major breakthrough in the development of IoT. Li-Fi networks, will provide a great connectivity on a higher bandwidth for the objects interconnected on the concept of IoT. Also. Bi-Directional technology gives a higher data rate up 40G Ethernet for a big data from multifarious devices of IoT [1].

IV. Characteristics of The IoT

The IoT represents a future technology that shows some fundamental characteristics as follows:

• **Intelligence**: IoT makes things smart and enhances life with the power of data collection, artificial intelligence algorithms, and networks. Although the popularity of smart technologies, intelligence in the IoT is only means of interaction between devices, while user and device interactions are achieved by usual input methods and graphical user interfaces [28]. In addition, artificial intelligence can help IoT devices to interact with humans and other objects intelligently and make autonomous decision[31].

• **Dynamic changes**: The state of devices change dynamically, e.g connected and/or disconnected as well as the context of devices including location and speed. Moreover, some new items joining the network while others leaving without determining network boundaries [14]. The IoT devices can dynamically adapt to changing situations based on their operating conditions.

• **Limited energy** : Most IoT devices such, wireless sensors, software, actuators, and computer device are small and lightweight with limited resources, so they are designed to work with minimal energy consumption [28].

• **Connectivity** : New technologies for networking, enables network accessibility and compatibility and specifically IoT networking, it also provides new market opportunities for IoT that can be created by the networking of smart things and applications also, provides the common ability to consume and produce data [14].

V. IoT A possible future applications

The IoT has the capability to connect everyday objects. There are a number of possible future applications. In this section, we present few of these applications and examines the challenges that have also been identified;

A. Smart cities

The smart cities in the future depend on networking devices and it can be accessed by everyone all over the country. In the future, the smart cities will completely depend on every connected device. The distributes sensors in houses, offices and cities make human's life more comfortable in various aspects. Also, the IoT can help in the design of smart homes e.g., energy consumption management, interaction with appliances, detecting emergencies, home safety and finding things easily, home security etc. This, as a result, will lead to the development of many smart cities with many smart features, which include: smart planning, smart ICT development and infrastructure, smart energy, smart buildings, smart governance [7]. Therefore, these devices have enabled the automation of home activities by the adoption of various embedded devices[32].

Figure 4. Smart Cities Aspects[7]

B. Smart Health

Smart health care is built around specialized devices such as sensors, surveillance systems and wearable all of which must be connected together. There are several options to choose from that serve the medical IoT sector, target specific health care solution needs and easily adapt across multiple technologies. The marketplace for e-health monitoring devices is presently characterized by precise solutions that are equally non-interoperable and designed based on various architectures and platforms[7]. A close attention that required to hospitalized patients whose physiological status should be monitored continuously can be constantly done by using IoT monitoring technologies. Sensors are being used to get comprehensive physiological data and use gateways and the cloud to examine and collect the information and then transfer the analyzed data wirelessly to caregivers for additional analysis and review. Moreover, hospitals will be equipped with smart flexible wearable embedded with RFID tags which will be given to the patients on arrivals, through which not just doctors but nurses will also be able to monitor heart rate, blood pressure, temperature and other conditions of patients inside or outside the premises of hospital [1].

C . Smart energy and electric grid

A smart grid is related to the information and control and developed to have a smart energy management. The IoT provides more information about the behaviors of electricity suppliers and consumers in an automated way to improve the energy efficiency [32]. It also provides consumers with smart management of energy consumption such as smart meters, smart appliances, renewable energy resources. A smart grid that integrates the data and information technologies (ICTs) to the power network will enable a real-time, two-way interaction between suppliers and consumers, creating more dynamic interaction on energy flow, which will help deliver power more efficiently and sustainably[29]. Such applications and functions will be founded on

networked intelligent applications and devices and electric grid infrastructure components, mostly grounded on IoT ideas.

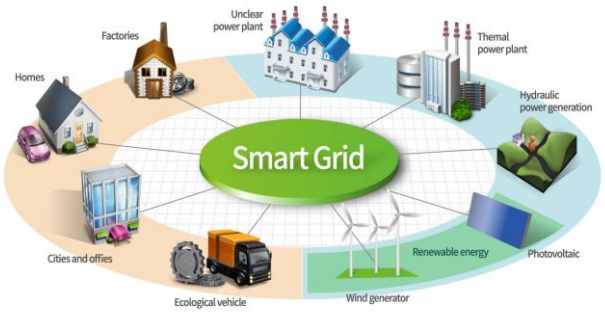


Figure 5. Smart grid applications

D. Smart homes, and Buildings

A smart home is a residence that uses internet-connected devices to enable the remote monitoring and management of appliances and systems, such as lighting and heating. Many companies are considering developing platforms that integrate the building automation with entertainment, healthcare monitoring, energy monitoring and wireless sensor monitoring in the home and building environments [31]. By the concept of the internet of things, homes and buildings may operate many devices and objects smartly, of the most interesting application of IoT in smart homes and buildings are smart lighting, smart environmental and media, air control and central heating, energy management and security.

VI. Security and Privacy needs

The IoT becomes a key element of the future internet, the need to provide adequate security for the IoT infrastructure is more important nowadays. One of the most difficult issues that face most of the new technologies is the security and privacy issues recently, therefore we have to look towards lightweight security solution to make IoT secure because this will put burden on device resources also, it can be harmful to the consumer's network. Many advanced security solution are required in several areas to make the IoT secure from attacks, thefts and many other security problems such as (Denial-of-Service) stack DoS/ Distributed Denial-of-Service (DDoS) attacks, compromised nodes, and malicious code hacking attacks, that because the IoT is susceptible to such attacks and will require specific techniques and mechanisms to ensure that transport, energy, city infrastructures cannot be disabled or subverted [33]. On the other hand, without security for IoT, any connected object, from refrigerators to manufacturing bots, can be hacked. Once hackers gain control, they can usurp the object's functionality and steal the user's digital data. Data encryption techniques is also very important in IoT solutions from threats based systems for enable a means of protection for data to be stored processed, without the data content being accessible to other parties [34]. Also from a legal point of view; including the influence of sites on the regulation of their privacy, the issue of information ownership in the collaborative clouds of the "things" data network and data anonymity can provide a privacy basis to protect.

VII. CONCLUSION

Internet of things is a technology which provides many applications to connect the things to things and human to things through the internet. The emerging idea of Internet of Things (IoT) is quickly finding its path throughout our modern life, aiming to enhance the quality of life by connecting various smart devices, technologies and applications. Generally, the IoT would allow for the automation of everything around us, exchange information and take intelligent decisions. All networks and technologies of communication are used in building the concept of the internet of things such technologies are mobile computing, RFID technology enables, wireless sensor networks, software, and embedded systems, in addition to many algorithms to get management processes, storing data, and security. This paper presented an overview of the IoT system. We have articulated research about layered architectures. The state-of-the-art and layered architecture of the IoT are discussed. IoT is also pitted against M2M to illustrate the similarities and difference between the technologies. In addition, IoT essential features and different communication technologies are presented. In addition to the IoT future applications directions are discussed. Finally, a review of security, privacy and protection of user networks.

References

[1] Farooq, M. Umar, et al. "A review on internet of things (IoT)." International journal of computer applications 113.1 (2015): 1-7.

[2] zzawi MA, Hassan R, Bakar KA. A review on Internet of Things (IoT) in healthcare. International Journal of Applied Engineering Research. 2016;11(20): PP: 10216-21.

[3] Acharjya, D. P., M. Kalaiselvi Geetha, and Sugata Sanyal, eds. "Internet of Things: novel advances and envisioned applications." (2017).

[4] Alhalafi N, Veeraraghavan P. Privacy and Security Challenges and Solutions in IOT: A review. InIOP conference series: Earth and environmental science 2019 Aug 1 (Vol. 322, No. 1, p. 012013). IOP Publishing.

[5] Mallela R, Nagendra P, Ramana K. Internet of Things–Future Internet Technologies, Elements and Applications. International Journal of Research and Analytical Reviews. 2018 Oct;5(4):960-7.

[6] Jing Q, Vasilakos AV, Wan J, Lu J, Qiu D. Security of the Internet of Things: perspectives and challenges. Wireless Networks. 2014 Nov;20(8):2481-501.

[7] Zeinab KA, Elmustafa SA. Internet of things applications, challenges and related future technologies. World Scientific News. 2017;2(67):126-48

[8] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide.

[9] Alam T. A reliable communication framework and its use in internet of things (IoT). CSEIT1835111| Received. 2018 May 1;10:450-6.

[10] Aljohani M, Alam T. Real time face detection in ad hoc network of android smart devices. In Advances in Computational Intelligence 2017 (pp. 245-255). Springer, Singapore.

[11] Bandyopadhyay D, Sen J. Internet of things: Applications and challenges in technology and standardization. Wireless personal communications. 2011 May 1;58(1):49-69.

[12] Gokhale, Pradyumna, Omkar Bhat, and Sagar Bhat. "Introduction to IOT." International Advanced Research Journal in Science, Engineering and Technology 5.1 (2018): 41-44.

[13] Granjal, Jorge, Edmundo Monteiro, and Jorge Sá Silva. "Network - layer security for the Internet of Things using TinyOS and BLIP." International Journal of Communication Systems 27.10 (2014): 1938-1963.

[14] Patel KK, Patel SM. IOT: definition, enabling technologies, application & future challenges. International journal of engineering science and computing. 2016 May;6(5).

[15] Abdulrahman, Lozan M., et al. "A state of art for smart gateways issues and modification." Asian Journal of Research in Computer Science (2021): 1-13.

[16] Khan R, Khan SU, Zaheer R, Khan S. Future internet: the internet of things architecture, possible applications and key challenges. In2012 10th international conference on frontiers of information technology 2012 Dec 17 (pp. 257-260). IEEE.

[17] Darwish, Dina. "Improved layered architecture for Internet of Things." Int. J. Comput. Acad. Res.(IJCAR) 4 (2015): 214-223.

[18] Burhan, Muhammad, et al. "IoT elements, layered architectures and security issues: A comprehensive survey." Sensors 18.9 (2018): 2796

[19] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems. 2013 Sep 1;29(7):1645-60.

[20] Kaur J, Kaur K. Internet of Things: A Review on Technologies, Architecture, Challenges, Applications, Future Trends. International Journal of Computer Network & Information Security. 2017 Apr 1;9(4).

[21] Albishi, Saad, et al. "Challenges and Solutions for Applications and Technologies in the Internet of Things." Procedia Computer Science 124 (2017): 608-614.

[22] Indu, Sunita Dixit. "Wireless sensor networks: Issues & challenges." International Journal of Computer Science and Mobile Computing (IJCSMC) 3 (2014): 681-85.

[23] Kitanov S, Monteiro E, Janevski T. 5G and the Fog—Survey of related technologies and research directions. Technical Conference (MELECON) 2016 Apr 18 (pp. 1-6). IEEE.

[24] Nguyen, Dinh C., et al. "6G Internet of Things: A comprehensive survey." IEEE Internet of Things Journal (2021).

[25] Verma, Pawan Kumar, et al. "Machine-to-Machine (M2M) communications: A survey." Journal of Network and Computer Applications 66 (2016): 83-105.

[26] Kubo. (2014). The research of IoT based on RFID technology. In 2014 7th international conference on intelligent computation technology and automation, Changsha, pp. 832–835.

[27] Bhatia, S., Chauhan, A., & Nigam, V. K. (2016). The Internet of Things: A survey on technology and trends. Information Systems Frontiers, 17, 261–274.

[28] Atlam, Hany Fathy, Robert Walters, and Gary Wills. "Internet of things: state-of-the-art, challenges, applications, and open issues." International Journal of Intelligent Computing Research (IJICR) (2018): 928-938.

[29] Al-Sarawi S, Anbar M, Alieyan K, Alzubaidi M. Internet of Things (IoT) communication protocols. In 2017 8th International conference on information technology (ICIT) 2017 May 17 (pp. 685-690). IEEE.

[30] Tiwari, Rajinder. "An overview of Internet of Things (IoT): From literature survey to application implementation perspective." International Research Journal of Engineering and Technology (2017): 575-582.

[31] Stojkoska, Biljana L. Risteska, and Kire V. Trivodaliev. "A review of Internet of Things for smart home: Challenges and solutions." Journal of Cleaner Production 140 (2017): 1454-1464

[32] Mahmood, Anzar, et al. "An overview of load management techniques in smart grid." International Journal of Energy Research 39.11 (2015): 1437-1450

[33] Shafiullah, G. M., et al. "Smart grid for a sustainable future." Smart Grid and Renewable Energy 4.1 (2013): 23-34.

[34] Bertino, Elisa. "Data Security and Privacy in the IoT." EDBT. Vol. 2016. 2016

IT Security Office for Private and Public Organizations: What, Why and How?

Ibrahim E. Lahmer
Computer Department, NEDB & NOC
i.lahmer210@gmail.com & ilahmer@noc.ly

***Abstract*— This paper, presents an IT Security Office proposal for any Libyan organization or institution that is planning for digitizing or having partially digitized their business model. The proposal is based on the need for any organization to secure; according to the international IT policy and standards, (1) the current condition of the use and the implementation of the Information Technology (IT) in the organization, and (2) the vision of the organization to adopt for example an electronic administrative and financial systems. This document is organized as following. First, the proposal introduces IT and its Security Office function and adoption. Then, the proposal specifies the administrative hierarchy of the office in the organization chart and why we need it. Third, it defines the working conditions of the office in such organizational chart. Fourth, the office duties and responsibilities are described. Finally, the proposal gives recommendations of how an organization could start working in this office and the benefits of implementing such recommendations.**

Keywords—IT, Security Office, Function, Organization, IT Operation, Standards, and Procedures.

I. INTRODUCTION

Information Technology (IT) and its role have become a vital and integral part of every organization function. The IT leads the organizations' businesses and strategies. For example, IT systems are responsible for organizing the administrative and financial processes of any organization or institution. IT systems should be implemented in a way that meets the needs of the organization, so that these systems contribute (and not limited) to the following benefits:

- Saving time and effort for both administrators and workers in any organization.
- Improving the production process and raising the quality.
- Getting rid of administrative and financial corruption.
- Assisting decision makers.

However, these benefits or the effectiveness use of the IT rely on the accuracy of the managed systems and information. The IT Security Office assures such accuracy. It provides support for creating policies and standards of implementing and using the Information Technology (IT), and it monitors and follows up the implementation, operation, and use of IT systems. This is to ensure (1) the protection and the continuity of the function of the information systems and infrastructure according to the international standards of information security, and (2) the protection of the Confidentiality, Integrity, and Availability (CIA) of the information and the data used and stored in the IT resources[3].

The presence of this office has become a major support for large institutions, companies, and international organizations, whether governmental, commercial or voluntary. There is a noticeable adoption of this office in these bodies. In 2011, the PricewaterhouseCoopers conducted a global study/survey to determine that to what extent the IT Security Office was adopted and used in these bodies. In their annual Information Security Report of the same year, the study showed that in 2006, 43% of about 200 of these bodies had an information security office, and in 2008 it reached to 56%. This percentage was doubled to reach 85% in 2009. The increase in the adoption and use of such office gives a significant sign of its importance[1][2].

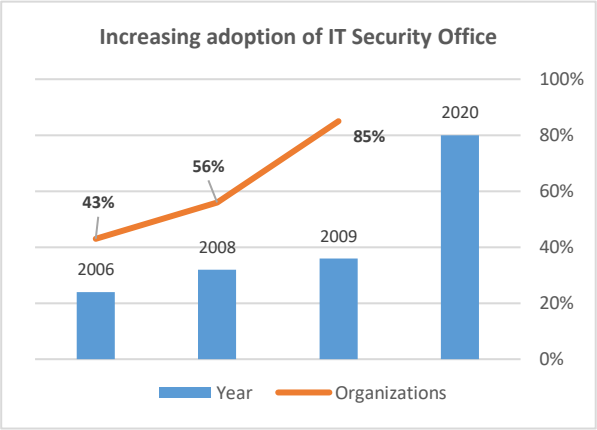


Fig. 1. Increasing adoption of IT Security Office.

We note that the increase in the adoption of this office doubled in three years between 2006 and 2009. In the time of writing this paper, it is more than ten years! If the same study is conducted now (i.e. 2022) the ratio could be larger and may reach 95% or more. Libyan organizations: public and/or private, should not be among those may fail to adopt this office or administration. This is because of the importance of its duties.

A. *Administrative Hierarchy of the IT Security Office*

According to the Global State of Information Security Survey (GSISS), companies are increasingly recognizing the importance of having a top-level manager or executive dedicated to security issues. Based on IDG's 2021 Security Priorities study, among more than 600 organizations surveyed in this study, almost two third of security managers/chiefs have a direct connection to the top management. As shown in Figure 2, 40% of them, the top security executive reporting to the CEO,

and 27% reporting to the Board of Directors (BoD). The study also showed the fact that more businesses are adding security executives to their leadership teams– 67% in 2021 versus 61% in 2020[4].

In order to ensure the performance of this office in any organization, it should be directly associated with the top management (e.g. Chairman of the Board of Directors) to achieve the following: (1) coordinating efforts between all the organization departments that use and implement the IT; (2) directly meet the needs of this office without delay; (3) submitting reports related to the IT security status of the organization to both the top management (Board of Directors) and the managers of the Information and Communication Technology (ICT) departments to take the appropriate action/implementation plan without any delay.

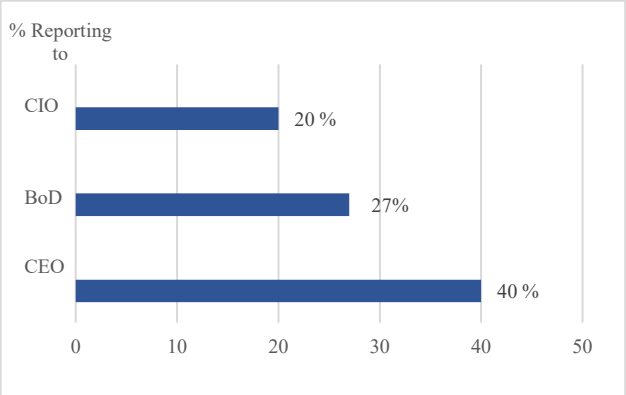


Fig. 2. IT Security Office Reporting to Top Management.

The important existence of such office with this administrative hierarchy in the organizational chart supports the need to (1) be more proactive in managing their information risk, and (2) face the greatest challenges, namely, the protection of their information systems and data. This protection has multiple aspects, and they include all the information resources of the organization that are distributed between different IT sectors and other organization departments. The presence of this office with this administrative hierarchy obligates and imposes on these departments the implementation and application of IT security regulations and standards that are needed for the protection of the data and its IT resources. However, the question raises is that what are the essential requirements or working conditions of such office to achieve its duties successfully?

B. Working Conditions (Requirements)

- Firstly; Working alongside with the ICT managers who manage the implementation and operation of the information and communication systems in the organization. This is to ensure that the ICT systems are implemented according to the ICT security policies, standards and procedures, so that reducing the vulnerabilities and preventing an ease of penetration of the ICT infrastructure and systems by their implementation.

- Secondly; Working alongside with other managers of the organization, where their employees use the IT resources to facilitate the tasks assigned to them. This is to ensure that the use of the IT is according to an acceptable use policy (AUP), educating the organization employees, and reducing the feasibility for penetrating the IT infrastructure and systems by their use.

- Thirdly; Knowing and following up of any decision or plan related to IT and communication projects, whether small or large projects, from IT departments or from other departments, in order to provide support from a standpoint of the IT security and its requirements. Here are two simple examples. First, when an organization decides to use a new IT technology or system which is based on the use of various web development tools (Frontend development tool, e.g. Angular.js) and internet browsers. These might pose security risks or threats [5]. When the IT Security Office is aware of the software developers needs to satisfy the business goals, the office can factor that into the decisions regarding the security policy and technologies such as implementing firewalls, network and database security mechanisms, or even assessing the security of the developed code, and enforcing other security mechanisms as needed. Second example is that when the organization decides to make a new data center for disaster recovery, the office can examine the IT security requirements of such data center. This includes, but not limited to, the IT physical security, the organization data and IT resources availability, and access monitoring and control standards of such data center [6].

II. GENERIC IT SECURITY OFFICE DUTIES

This section describes in brief the main functions of the IT Security office which are presented in:

- Overseeing all the IT security policies, standards, and procedures for the organization. This includes the IT policy and standards for both implementing and using the IT systems and infrastructure.

- Following up the software and technical solutions that are related to the performance and the implementation of IT security procedures.

- Following up the applications and software solutions that are designed and developed by the IT department team of the organization, if any, or by a third party company and provided to organization departments to facilitate their functions (i.e. daily tasks). This is to ensure the use of system development standards and information security mechanisms in these applications and software solutions, protecting their availability, and protecting them against an easy penetration by reducing the weak points in these software solutions. This is because weak points allow misuse of these solutions and hackers to easily access sensitive information and the data of the organization [19].

- Following up the latest technology trends in the IT security, including new security software and tools as well as the best knowledge of how the IT hackers and intruders are behaving.

III. DETAILED IT SECURITY OFFICE DUTIES AND RESPONSIBILITIES

In details, the office duties can be categorized in five main components; IT Policy and Standards, Management Support, IT Monitoring and Testing, IT Incident and Risk Management, IT Audit and Follow up Management [8]-[13].

A. IT Policy and Standards

This office provides direction and practical setting for IT Organization security policies and procedures to protect the IT critical resources and services which includes data, information and IT appliances. This could be achieved according to the following:

- Creating the IT usage policy to avoid the misuse of the IT in the organization.
- Creating the IT policy of the security standards and procedures. This is to reduce the weaknesses and prevent the ease of penetration of the IT infrastructure as a result of setting and implementing the IT for the organization.
- Reviewing, from time to another, the standards and procedures used in setting up and implementing the IT, and developing a vision to improve these standards and procedures when necessary as the used IT is evolved.
- Working with the organization legal affairs department to review and document the IT policy and regulations in term of legal aspects.
- Working with the top management, e.g. Board of Directors, to ratify these IT policies and regulations and to enforce their implementations.
- Working with other organization departments and managements to support the implementation of these IT policy and regulations.
- Working with the organization legal affairs department to study and review the General Data Protection Regulation (GDPR). This is to include the necessary terms in the IT usage policy and security standards regulations that protect the user and the implementer of the IT in the organization.
- Working with the organization legal affairs department to evaluate the existing and emerging IT-security related laws, regulations, and policies to be complied with.

B. Management Support

This office assists the ICT managers and supervisors with ICT security administration, implementation, and management. This includes:

- Implementing the IT security policy, standards and procedures.
- Ensuring the existence of the IT security tools from appliances and software in all of the IT infrastructure and data centers. This is to enforce the implementation of the IT policy, standards and procedures that are ratified.
- Looking and studying the latest technology trends in the IT security to specify what is needed to improve the safety and security of the organization's IT resources, information and data.
- Testing and evaluating the chosen IT security technologies, advising on security related technology projects, and aiding in the management of security technology, with special emphasis on mission critical IT resources.
- Raising the awareness and the culture of information security for the employees of the organization.

C. IT Monitoring and Testing

This office conducts the IT Health Check and IT security testing and monitoring for the organization. This is to support the implementation and the optimal operations for the IT policies and practices and to identify any IT security threats or risks. This includes:

- Monitoring computers and network resources for any suspicious activity through the IT infrastructure.
- Testing/assessing information resources, internally and externally, for any security breaches or vulnerabilities in the IT infrastructure.
- Conducting security scanning of the computers and servers that host and run the organization data and applications. This is to reveal any foreign programs that cause direct or indirect harm to the application and the data of the organization.
- Testing the IT windows and web-enabled services of the organization to discover any weakness in the developed applications of such services.
- Monitoring and logging events from various IT resources used and record theses logs in Data Logs.
- Conducting IT security analysis on the Data logs which are recorded to specify the performance of the IT resources and to identify any IT security threats.

D. IT Incident and Risk Management

This office manages and oversees the incident response, investigation, and reporting regarding the IT security incidence and events. It also, maintains an information security risk management program to evaluate threats and vulnerabilities that have been identified via monitoring and testing functions, and it assures creation of appropriate remediation plans. This includes the following:

- Performing network intrusion detection and conducting forensic and administrative investigations related to such incidence and events.
- Receiving and processing IT security incident complaints, and overseeing recovery, and restoration for IT security-related events.
- Taking the permitted and required actions to protect the IT resources of the organization in coordination with the related departments such as ICT and service departments.
- Providing support in assessing IT security risks, creating and monitoring IT security plans, and aiding IT disaster recovery planning.

E. IT Audit and Follow up Management

The office audits (i.e. examines and evaluates) the IT procedures and operations to ensure that they are complaint with both the IT security policy and standards and the needs of the organization. This includes the audit and the follow-up of the following:

- Controls of the IT implementation.

- Controls of the maintenance and analysis for the entire security audit logs.

- Controls of the use and monitor of the administrative privileged IT accounts.

- Controls of application software security.

- Controls of data back-up operations.

- Controls of security of wireless communication.

- Controls of limitation and monitor of ports, protocols and services.

- Controls of continuous vulnerability testing and remediation.

- Controls of secure configurations of network devices such as firewalls, routers and switches.

- Controls of IT physical security.

IT security auditing determines whether the IT security controls mentioned above ensuring the following:

- ✓ Protecting the confidentiality, integrity and availability (CIA) of the organization data, information, and the IT resources; such as software applications, servers and PCs resources.
- ✓ Aligned with the organization's business overall goals.

IV. RECOMMENDATIONS TO START WITH

This office should be established due to both the importance and the size of its duties. Carrying out the tasks of this office, while the existence of ICT departments and other departments that use (and may implement) the IT, is a complex task. However, it is not difficult to accomplish. In order to avoid the failure of the work of this office when it is established, and to start its main tasks/duties smoothly and working in steady pace and effectively, the Office should start by doing the following preliminary tasks.

A. Preliminary Task: Documented Technical Study and Proposals

• Conducting a comprehensive technical study to document the existing IT standards used by the ICT departments that implement the ICT systems in the organization. The results of this study are to identify any problems facing these departments in implementing their work according to the international standards for information security.

• Throughout this study, the existing criteria (standards) used in implementing the IT is/are defined, and evaluated from the IT security standpoint. Based on the outcome of this evaluation, the existing (used) standards are categorized into two types: (i) standards that are consistent with international standards of information security, and this type can be maintained, and (ii) standards that are inconsistent with international standards of information security which may lead to (1) compromise the confidentiality and the integrity of the organization data and its sensitive information, and (2) affect the functionality of the organization when using the IT to run its business; specially when there is a potential to implement electronic administrative and financial system.

• Providing practical and technical proposals for implementing solutions to solve the problems facing ICT departments in implementing and applying the IT according to the international standards of information security.

• To typically implement these proposals, they are formulated into two types:

- The first type of proposals deals with the problems/issues that can be resolved directly without pre-requirements or needs to apply the solutions of these problems. In this type of proposals, the solutions are written in the form of scripts and then interpreted into executive regulations (standards) for their application.
- The second type of proposals is concerned with problems/issues that cannot be resolved directly, but needs to be solved using small projects supported by regulations. This is because of the requirements that must be provided before implementing these solutions. In this type of proposals, the solutions are written in the form of texts and then interpreted into executive regulations for their application. These projects identify the requirements, the stages of each project, the estimated budget for each one and its completion as well as the aspects of budget expenditure.

• Periodic reports shall be submitted to the one who is entitled[4] to follow up the work of this study, i.e. the one who is responsible to follow up the preliminary work of this office.

• Whenever the regulations that are related to the duties of the IT Security Office are implemented, the performance of these regulations MUST be evaluated regularly, and periodic reports are presented.

B. The Benefits of This Study

International IT security frameworks and standards [8][19] can be used to direct this study to ensure that the study aligns with the information system auditing mechanisms and its realization of the following potential and promised benefits:

1. Start working quickly in the duties of the office, as some of the tasks achieved during this study are part of the office main tasks.
2. Providing support to improve the performance of the IT department. This is done via defining the current duties/roles of IT managers, engineers, and technicians, and determining the shortages in the IT roles that are needed according to the IT security standards. A list of recommendations and suggestions are advised to both the one who is entitled to follow up the work of this office and the IT departments. This is to establish and reassure the IT roles according to the international standards of information security, so that the organization can implement the IT technologies to facilitate its function according to the international standards. This list of recommendations and suggestions is part of IT change management process. In other words, this study can contribute in defining the chains of responsibility and authority.

3. Start working with the IT governance process. This is because defining the chains of responsibility in term of IT security is one of the main key components of IT governance process. As described by COBIT framework and IBM IT Governance Approach, it "is used to define the chains of responsibility, authority, and communication to empower people, as well as to define the measurement and control mechanisms to enable people to carry out their roles and responsibilities"[16]-[19].
4. Identifying the first component of a systematic approach used in Information Security Management System (ISO 27001) which consists of three main components; People, Processes, and Technology that help to protect and manage all the organization Data and IT resources through IT Incident and Risk Management.
5. By defining the current condition of the IT policy and standards used in the implementation of the IT projects in the organization, the study could find out the reasons and issues that hinders the best practice of implementing and using IT in the organization. Thus, the study could support a detailed plan/s for the digital transformation project in term of IT Security need based on the organization vision .
6. Last but not least, the study can identify the priorities/preferences that must be achieved from the functions of this office (in Section III (1 to 5)) and not included in the work of this study. This relies on the needs, the importance of the work to be done, and the budget available to the office and the ICT departments, as well as the top management approvals. Note that the implementation of the duties of this office will also rely on (i) identifying the components and the requirements of this office; and (ii) developing plans that must be documented and reviewed before implementing them in coordination with the related departments.

C. The Study Needs (Requirements/Working Conditions)

1. The consent and collaboration of the IT and ICT managers is required to conduct this study.
2. The office chooses the people who are needed from the IT or other organization departments in coordination with the managers of these departments, to assist in accomplishing this study.
3. Collaboration of IT engineers, developers and technicians to complete this study, so that fist block or component of the IT security office could be made, i.e. roles and responsibilities and we start work according to the IT standard.
4. Non-intervention of others in the work of this study is essential, except when it is needed.
5. Transparency in the study is important, so as to be able to maximize the performance in implementing and using the IT at the organization.
6. Confidentiality of some of the outcomes of this study is also important, the result or the outcomes that may affect the privacy and the security of the organization data or some of the staff of the IT or other departments in the organization.
7. Collaborate with some international organizations of IT security and standards when needed.

D. Additional Requirements to run the Office

The following additional requirements may be needed to get use of the outcomes of the study in achieving the office duties successfully

- Select the consultants from the competent companies, and sign supporting service agreements when necessary. This is to carry out the duties of this office according with the international IT security policy and standards.
- The office chooses competent engineers and technicians, if any, from the IT departments to implement and follow up the work of this office in coordination with ICT managers, and from outside to perform some other duties as needed and described above.
- Assign budget to this office to perform its duties and to attend both training courses and workshops that are related to its duties.
- Non-intervention of others in the work of this office unless it is requested.
- The office needs to be independently represented in the IT security's needs, goals and vision for the organization and not be buried too deep into an operational capacity of the IT security.

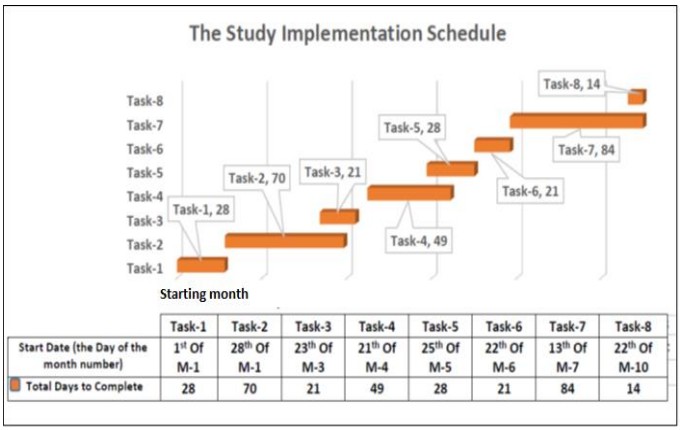
E. The Study Implementation Tasks and Schedule

In any organization, the scheme of the implementation tasks includes questionnaire to be answered, observation on IT operations and interview of the IT admins and supervisors. This scheme should be supported by Information System Auditing guidance and procedures. This section presents the study implementation tasks and schedule. The study implementation schedule presented with a part time job IT system auditor. Note that the starting date of such study is not specified, as it is based on the approval of this proposal in any organization. The period assigned to this study is subjected to increase and decrease depending on the existing working condition of the IT team of the organization. However, during this study period, it is feasible to start implementing some proposals that solve the existing issues of not implementing and using the IT according to the international standards of information security.

Table-1: Tasks Details

Tasks	Description
Task-1	Preparing for the study by understanding the existing of the organization ICT departments and identifying a strategy to conduct the study.
Task-2	Collecting the data and information regarding the used IT policy and standards.
Task-3	Documenting the collected data and information.
Task-4	Evaluating the collected and documented data/information of the used IT policy and standards based on the international IT policy and standards for information security.
Task-5	Identifying the problems and issues that hinder the typical use of IT policy and standards.
Task-6	Categorizing the existing IT policy and standards used in the IT department.

Task-7	Creating and classifying technical and documented proposals for solutions to address the existing issues of using IT policy and standards.
Task-8	Reviewing the proposals to be implemented.



	Task-1	Task-2	Task-3	Task-4	Task-5	Task-6	Task-7	Task-8
Start Date (the Day of the month number)	1st Of M-1	28th Of M-1	23th Of M-3	21th Of M-4	25th Of M-5	22th Of M-6	13th Of M-7	22th Of M-10
Total Days to Complete	28	70	21	49	28	21	84	14

Fig. 3. The study implementation schedule.

V. Discussion and Future Work

This paper has defined the IT Security office and why it is needed. It has also proposed a method of how any Libyan organization could start its implementation. The IT Security Office focuses exclusively on delivering the highest levels of IT security and assurance which could be implemented in five functions; IT Policy and Standards, Management Support, IT Monitoring and Testing, IT Incident and Risk Management, IT Audit and Follow up Management. Studying the AS-IS of the existing IT service management can identify any problems and gabs in the IT implementation and use. The IT standards, policies and procedures are the measurements in such study. The expected output of the study is to assess and measure the performance of IT processes within the organization, and for the best performing organization, the output of the study should divide the management and operational responsibilities between IT operations and the information security and assurance functions. However, a study and/or more research is needed to define the IT Security Office organizational chart/team and the roles and responsibilities of the information security personnel.

References

[1] Boris Vaynberg, "How has the CISO function changed through the years?", Feb 26, 2019, https://www.mimecast.com/blog/ciso-at-25/ *Retrieved Dec 2021,* and "Annual Information Security Survey", *PricewaterhouseCoopers pwc. Retrieved Dec 2012* https://www.pwc.com/gx/en/issues/cybersecurity.html.

[2] Mike Maali and Thack skidmore, "Real-time compliance management Report", *PricewaterhouseCoopers pwc. Retrieved Dec 2021*, https://www.pwc.com/us/en/risk-assurance/state-of-compliance-study/assets/pwc-2018-state-of-compliance.pdf.

[3] ISO/IEC 27000:2018 "Information technology and Security techniques, ISMS-Overview and vocabulary, 2018 2018-02, Edition : 5 Number of pages : 27, Technical Committee : ISO/IEC JTC 1/SC 27,Information security, cybersecurity and privacy protection, https://www.iso.org/standard/73906.html

[4] International Data Group (IDG) Research Team, "A Crisis Of Confidence Pushes New Security Initiatives", 2021 Executive Summary, IDG Security Priorities Study IDG COMMUNICATIONS, INC. https://www.idg.com/tools-for-marketers/research-security-priorities/

[5] William C. Barker & et. al. June 2020, "Securing Web Transactions, TLS Server Certificate Management", NIST SPECIAL PUBLICATION 1800-16, *Retrieved Feb 2022* https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1800-16.pdf

[6] Security Policy, NISSA Publication, National Information Security & Safety Authority, *Retrieved Feb 2022* https://nissa.gov.ly/en/main-services/policies/

[7] Mark Simos, Aug. 2020 "How to organize your security team: The evolution of cybersecurity roles and responsibilities" last retrieved Feb 2022, https://www.microsoft.com/security/blog/2020/08/06/organize-security-team-evolution-cybersecurity-roles-responsibilities/

[8] ISACA Publication, Certified Information Security Auditor (CISA), Review Manaual 27th Edition, updated 2019 Job Practice, ISBN 978-1-60420-767-5, CISA® Review Manual 27th Edition.

[9] CISA Publication, 2021, "Cybersecurity and Infrastructure Security Agency (CISA)", National Cybersecurity and Communications Integration Center's (NCCIC), Computer Emergency Response/Readiness Team of United Sate (US-CERT), last retrieved Dec 2021,), https://www.us-cert.gov/about-us, https://www.cisa.gov/about-cisa.

[10] SANS Publications, IT security recommendations and models, SANS Institute of advanced technologies for information management and security, last retrieved Feb 202, https://www.sans.org/.

[11] NIST Publications and Resources, The National Institute of Standards & Technology (NIST), and National Volunerabilty Database (NVD), last visited Feb 2022, (https://nvd.nist.gov/general , https://www.nist.gov/), .

[12] IT Security Office Duties, Guides and Policies of Standford Univerosity, Northwestern University, and the University of Texas, *Retrieved Nov. 2021, https://uit.stanford.edu/security, https://www.it.northwestern.edu/about/departments/iso/index.html, and https://www.utexas.edu/*

[13] Julia H. Allen, & et. al. "Structuring the Chief Information Security Officer Organization", Sep. 2015, SEI Technical Report for roles and duties of the IT Security Office, Carnegie Mellon® and CERT , *Retrieved Jan 2022, https://resources.sei.cmu.edu/asset_files/TechnicalNote/2015_004_001_446198.pdf*

[14] US Department of HomeLand Security (DHS) Publication, Sep, 2013, "Information System Security Officer Guide" last retreived Feb 2022 https://www.dhs.gov/sites/default/files/publications/Information%20System%20Security%20Officer%20%28ISSO%29%20Guide.pdf ,

[15] InfoSec Publications, Online Repository for Information Security (InfoSec) Reports, *last visit Feb 2022* https://informationsecurity.report/

[16] ITPC, IT Policy Compliance Group, Feb 2010, "Best Practice for Managing Information Security Report", *Retrieved Oct. 2020* http://eval.symantec.com/mktginfo/enterprise/other_resources/b-best_practices_for_managing_information_security-february_2010_OR_2876547.en-us.pdf

[17] LOGIIC Group, LOGIIC public reports "Improving cybersecurity in the oil and gas sector", LOGIIC: Linking the Oil and Gas Industry to Improve Cybersecurity, *Retrieved Aug.. 2021* . https://www.automationfederation.org/Logiic/LogiicProjects

[18] Lynn Mueller, & et. al. 2008, "IBM IT Governance Approach, Business Performance through IT Execution, 1st edition *Retrieved Jan. 2021, http://www.redbooks.ibm.com/*.

[19] John Lainhart & et. al., 2019 "*Cobit® 2019 Framework: Introduction & Methodology" COBIT:* Control Objectives for Information and Related Technology, ISACA Publication *Retrieved July. 2021 http://www.isaca.org/*

[20] ISO/IEC Publication, 27001Standards,Information Security Management System (ISMS) last visited in Feb 2022, https://www.iso.org/

الطاقة المتجددة والنمو الاقتصادي- دراسة تطبيقية قياسية على دول عربية مختارة، بالتركيز على دولة ليبيا للفترة (1990-2020)

د. داليا علي منصور، د. ليلى التهامي مرغم
{da.mansour,l.morghem}@uot.edu.ly
كلية الاقتصاد والعلوم السياسية
قسم التخطيط المالي
جامعة طرابلس – دولة ليبيا

الملخص يمثل قطاع الطاقة المتمثل في النفط والغاز الطبيعي المصدر الرئيسي لتمويل الخزانة العامة للاقتصاد الليبي، إذ يؤدي نضوب هذا المصدر إلى حدوث أزمة اقتصادية حادة في حالة عدم وجود مصادر تمويلية بديلة. حيث تتوفر في ليبيا موارد متعددة للطاقة المتجددة يمكن استغلالها ومن أهمها الطاقة الشمسية حيث يصل المتوسط السنوي للإشعاع الشمسي إلى 7 كيلوواط / ساعة على المتر المربع في اليوم، وأما عن طاقة الرياح يبلغ المتوسط السنوي لسرعات الرياح في ليبيا ما بين 5-10(متر / ث) . على الرغم من حدوث نمو سريع في مجال الطاقة المتجددة حول العالم، إلا أن ليبيا لم تستخدم إمكاناتها المتاحة في الطاقة المتجددة. ولم تستطيع الوصول لإستخدام 7% في سنة 2020، والتي تمثل مقدار مساهمة الطاقة المنتجة بالطاقات المتجددة من اجمالي الطاقة الكهربائية للمنظومة الوطنية كما كان مستهدفاً (حسب المخطط الوطني لتطوير الطاقات المتجددة).

تهدف هذه الدراسة إلى فحص العلاقة بين استهلاك الطاقة المستمد من مصادر الطاقة المتجددة والطاقة غير المتجددة والنمو الاقتصادي معبراً عنه بالناتج المحلي الإجمالي. وتتضمن مجموعة البيانات المستخدمة للدول العربية الواقعة في شمال القارة الأفريقية، للفترة من 1990 إلى 2020، بالتركيز على الدولة الليبية، وباستخدام البيانات المقطعية عبر الزمن غير المتوازنة (Unbalanced Panel Data)، وبتوظيف مناهج الاقتصاد القياسي واستخدام دالة الإنتاج (Production Function). وبتطبيق منهجية (ARDL) اختبرت الدراسة فرضية وجود علاقة ذات دلالة إحصائية بين الطاقة المتجددة والنمو الاقتصادي، وتوصلت إلى وجود علاقة طردية ذات دلالة إحصائية بين الطاقة المتجددة والنمو الاقتصادي في الأجل الطويل وغياب هذه العلاقة في الأجل القصير، وأوصت الدراسة بالاستغلال الأمثل لمناخ وموقع ليبيا الجغرافي في انتاج الطاقة النظيفة، والاستفادة من الدول الرائدة في مجال الطاقة المتجددة .

الكلمات الافتتاحية: النمو الاقتصادي، الطاقة المتجددة، الطاقة غير المتجددة، اجمالي التكوين الرأسمالي، القوى العاملة.

1. المقدمة

شهدت العقود الأخيرة تزايد الاهتمام بمصادر الطاقة التقليدية في العديد من اقتصادات الدول العربية، لما تشكله عوائد هذا المصدر من دخل كبير في جانب إيرادات هذه الدول حتى أصبحت من الدول ذات الاقتصادات الريعية. ليبيا إحدى هذه الدول التي تعتمد بشكل كبير على هذا المصدر حتى بات استقرار الوضع الاقتصادي لديها منوط باستقرار أسعار النفط وعوائده.

أصبحت الحاجة ملحة للبحث عن استخدام وتطوير مصادر بديلة للطاقة، إذ أن تزايد استهلاك الطاقة التقليدية بشكل تصاعدي أدى إلى استنزاف الموارد الطبيعية وتزايد نسبة التلوث البيئي.

قامت الدول مؤخرا بزيادة الاهتمام بالطاقات المتجددة لأن لها مستقبل واعد ضمن الخريطة العالمية وخصوصا أن منطقة الشرق الأوسط وشمال أفريقيا تمتلك مصادر الطاقات المتجددة تقدر بحوالي نصف امكانات العالم، بالإضافة إلى أنها تتمتع بجو مشمس للحصول على الطاقة الشمسية وطاقتي الرياح والمياه، وبالتالي استخدام الطاقة المتجددة يعزز النمو الاقتصادي ويقود إلى التنمية المستدامة ويساهم في الحد من التلوث البيئي من خلال تعزيز الاقتصاد الأخضر. وتأسيساً على ذلك لابد من بيان دور الدول العربية في التحول إلى الاقتصاد الاخضر بالتركيز على دور الطاقة المتجددة في دعم الاقتصاد الوطني في ليبيا ودول العينة.

تكمن مشكلة الدراسة في زيادة اعتماد الدول وخاصة النامية منها على الوقود الأحفوري مما أدى إلى زيادة تلوث البيئة ونضوب مصدر الطاقة، الأمر الذي يجعل من الأهمية بمكان تحديد الدور التي يلعبه التحول إلى الطاقة المتجددة في النمو الاقتصادي. تهدف هذه الدراسة إلى تحديد العلاقة بين الطاقة المتجددة والنمو الاقتصادي ممثلاً بالناتج المحلي الإجمالي للدول الواقعة بالشمال الأفريقي مع التركيز على حالة الدولة الليبية، عن طريق استخدام النموذج القياسي المناسب ،وقد تم اختيار عينة الدراسة للتشابه الكبير بينهم من حيث الجغرافيا والمناخ وتنوع المصادر الطبيعية في كلٌ منهم، ويقصد بالطاقة المتجددة (الشمس والرياح وطاقتي المد والجزر).

تبرز أهمية الدراسة من أهمية موضوع التلوث البيئي الذي أصبح يتصدر جداول أعمال المنظمات الدولية والحكومية على حد سواء. حيث أن البحث في موضوع الطاقة المتجددة وعلاقتها بالنمو الاقتصادي والبيئة من شأنه بيان دور هذا القطاع في تعزيز ورفع الناتج المحلي الاجمالي لدول العينة إضافة إلى الطاقة غير المتجددة والعمالة ورأس المال المادي، إذ تشكل كل تلك المتغيرات عناصر العملية الإنتاجية. اعتمدت الدراسة على استخدام دالة الإنتاج (كوب – دوغلاس) باستخدام المنهج الوصفي التحليلي والمنهج الاحصائي في الوصول إلى التحقق من فرضية الدراسة التي تقضي بوجود علاقة ذات دلالة معنوية إحصائية بين عناصر الإنتاج والنمو الاقتصادي، وتغطي الدراسة الفترة من 1990-2020، وتضم العينة الدول (ليبيا – مصر – تونس – الجزائر – المغرب).

تأتي هذه الدراسة على النحو التالي: فيما يلي المقدمة تستعرض الدراسة الأدبيات في هذا المجال وما توصلت إليه الدراسات السابقة من نتائج، ثم في القسم الثالث تقوم الدراسة بعرض مفاهيم عامة حول الطاقة المتجددة، وحالة الطاقة المتجددة في الدول العربية وبيان المؤسسات والتشريعات المنظمة للطاقات المتجددة بدول العينة، وفي القسم الرابع تقوم الدراسة بتوظيف دالة الإنتاج في تقدير النموذج القياسي لتحديد العلاقة بين النمو الاقتصادي وعناصر الإنتاج المختارة لجانب الطاقات المتجددة، ويحتوي هذا القسم أيضاً على المنهجية المتبعة في التقدير، واختبار فرضية الدراسة وعرض النتائج، وأخيراً تأتي التوصيات التي تقترحها الدراسات بناء على النتائج التي توصلت إليها.

2. الإطار النظري

تعتبر منطقة شمال افريقيا من أكثر الاسواق أهمية بسبب وفرة مواردها الطبيعية، وحيث تقوم الطاقة المتجددة بدور رئيسي في النمو الاقتصادي مما يساعد على تحقيق التنمية الاقتصادية في أغلب اقتصادات العالم، وهو ما تؤيده الدراسات التطبيقية السابقة في تحديد العلاقة بين الطاقة المتجددة والنمو الاقتصادي، وفي دراسة للتعرف على أهم مصادر الطاقة

في ليبيا جاءت دراسة (عاشور، 2012)، لتبين وجود خلل في الاقتصاد الليبي يتمثل في افتقاره للتنوع في مصادر الدخل، حيث تعتمد ليبيا وبشكل كبير على قطاع النفط والغاز الطبيعي، الأمر الذي أثر سلباً على الجهود الرامية لتحقيق التنمية الاقتصادية، وخلصت الدراسة إلى ضرورة تطوير استخدامات مصادر الطاقة المتجددة ووضع خطة عمل لتغيير الأنماط غير المستدامة لإنتاج واستهلاك الطاقة. وعن كفاءة استغلال الموارد الطبيعية في الدول العربية، قامت دراسة (زهير، 2015) التي استخدم دالة الإنتاج في تحليل مصادر النمو الاقتصادي ومدى كفاءة استغلال الموارد الطبيعية بناء على قيمة إنتاجية العامل الكلية بالمقارنة بين دولتين عربيتين (الجزائر وقطر) كعينة من دول العالم العربي، توصلت الدراسة إلى اعتماد النمو الاقتصادي على الطاقة وتفاوت تكاليف النمو بين البلدين من حيث استنزاف الموارد الطبيعية.

يعد النمو الاقتصادي إحدى الأركان الأساسية للنمو والتنمية التي تسعى إليها الدول، وأن هذا النمو حسب النظرية التقليدية يحركه عاملين أساسيين وهما رأس المال المادي ورأس المال البشري، وقد أعتمد (بوعتلي) في دراسته سنة 2019 عليهما كمحركات للنمو الاقتصادي، وتم اختيار نوع من الدوال غير الخطية في تحديد توليفة الانتاج وكشفت الدراسة عن مدى فاعلية استخدام عناصر الانتاج وباستخدام دالة كوب دوغلاس الشهيرة - وقد ظهرت هذه النظرية بعد ظهور النظرية التقليدية للنمو لأدم سميث وروبرت مالتس ونظرية هارود للنمو فهي تساعد في معرفة كيف ولماذا يتم النمو؟ - وُجد أن هناك تأثير ايجابي لاستهلاك الطاقات المتجددة على النمو الاقتصادي بدول المغرب العربي خلال الفترة 1995-2014 لأنه يساعد في تحقيق احتياجات المناطق النامية من الطاقة، وأيدت دراسة (الحسنين) هذه النتيجة، حيث توصلت إلى وجود أثر مهم جداً لاستهلاك الطاقة المتجددة على النمو الاقتصادي في المغرب خلال الفترة من 1971 – 2015 (الحسنين،2021).

أوضحت دراسة (ثابت، 2017) أهمية الطاقة المتجددة ودورها كبديل تنموي من شأنه أن يقلل من المخاوف البيئية المتعلقة بالطاقة وتساهم في التنويع الاقتصادي وتخلق فرص عمل واسعة والتي من شأنها المساهمة في تطوير اقتصاد مستدام يحافظ على البيئة ، بالإضافة إلى تطوير رأسمال المال المادي ودورها في تحقيق النمو الاقتصادي وتسريع عجلة التنمية، بالإضافة إلى المحافظة على البيئة وتقليل التكاليف وتساهم في الحفاظ على الموارد المالية والمادية من الهدر، وهو الأمر الذي أيدته دراسةQinghua, Fu وآخرون سنة 2021 بإثبات وجود علاقة ثنائية الاتجاه بين استخدام الطاقة المتجددة والاقتصاد من خلال الزيادة في الناتج المحلي الإجمالي، ووجود علاقة سببية أحادية الاتجاه بين استخدام الطاقة المتجددة وانبعاثات ثاني أكسيد الكربون. كما وجدت دراسة حسن أن العلاقة بين استهلاك الطاقة والنمو الاقتصادي ذات دلالة احصائية في الأجل الطويل والقصير، وقد قامت هذه الدراسة في مصر عن الفترة 1970 -2020 واعتمدت على بيانات متوسط استهلاك الفرد من الطاقة الاولية ومتوسط نصيب الفرد من الناتج المحلي الاجمالي (حسن، 2020)، وتطرقت الدراسة للنظريات المفسرة للنمو الاقتصادي مثل النظرية الكلاسيكية والتي تعبر عن النمو الاقتصادي بدالة الانتاج (العمل – رأس المال – الموارد المتاحة – التقدم الفني)، والنظرية الكينزية والتي تقوم على نموذج Dommar Harrod وتعتمد على الاستثمار والنظرية النيوكلاسيكية وتعتمد على عوامل الانتاج ويعتبر التقدم التكنولوجي مصدرا أساسيا للنمو الإقتصادي وفقا لنموذج Solow (حسن، 2020) .

3.مفاهيم عامة حول الطاقة المتجددة

مفهوم الطاقة المتجددة: تُعرف الطاقة المتجددة على أنها الطاقة المكتسبة من عمليات طبيعية تتجدد باستمرار وبالتالي فهي عبارة عن مصادر طبيعية دائمة وغير نابضة ومتوفرة في الطبيعة سواء أكانت محدودة أو غير محدودة، ولكنها متجددة باستمرار وهي نظيفة لا ينتج عن استخدامها تلوث بيئي نسبيا، ومن أهم هذه المصادر الطاقة الشمسية التي تعتبر في الأصل هي الطاقة الرئيسية في تكون مصادر الطاقة وكذلك طاقة الرياح وطاقة المد والجزر والأمواج (راتول و مداحي 2012، 140). وتعرفها وكالة الطاقة الدولية (IEA) على أنها "طاقة مشتقة من العمليات الطبيعية التي يتم تجديدها بمعدل أسرع من استهلاكها"، وتذكر الطاقة الشمسية وطاقة الرياح والطاقة الحرارية الأرضية والطاقة المائية والطاقة الكتلة الحيوية كأمثلة على الطاقة المتجددة.

وقد أصبح استكشاف الطاقة المتجددة (Renewable energy) وتطويرها والطلب عليها في اتجاه تصاعدي، لعوامل عديدة مثل التأثير البيئي وأسعار النفط المتقلبة الأمر الذي ترتب عليه دفع الدول للتوجه نحو زيادة توليد الطاقة المتجددة، حيث تضاعف استهلاك الطاقة المتجددة في الدول الكبرى إلى الضعف خاصة الطاقة الشمسية وطاقة الرياح وهما أكثر مصادر الطاقة المتجددة فعالية وأقلها تكلفة وأيضا من أقلها تأثيرا على البيئة.

1.3 مصادر الطاقة المتجددة:

تطورت مصادر الطاقة التي ابتكرها الانسان بداية من الاعتماد على القوة العضلية ثم الاعتماد على الحيوانات وقوة الرياح والمياه في تحريك السفن، ومن ثم الاعتماد على الفحم وأخيرا أكتشف النفط والغاز الطبيعي، وأصبحت مصادر الطاقة متنوعة من مصادر ناضبة ومتجددة ودائمة (بوعتلي 2019، 15).

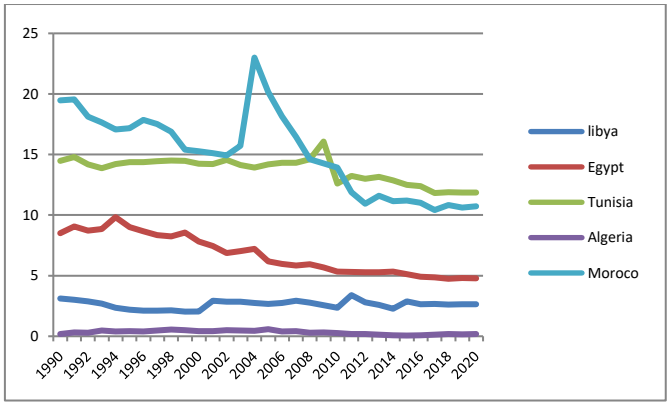
- **الطاقة الشمسية** هي الطاقة الشمسية الضوئية (باستخدام الضوء من الشمس) والحرارية الشمسية (باستخدام حرارة الشمس) وتتميز الدول العربية بأعلى إشعاع شمسي في العالم حيث تصل كثافة الاشعاع الشمسي 1000 وات / متر في منتصف النهار، ولهذا تتميز الدول العربية بتوافر الطاقة الشمسية بمعدلات تزيد عن مناطق العالم الأخرى كما هو الحال في دول العينة.
- **طاقة الرياح** هي الطاقة التي يتم الحصول عليها من الرياح وتستخدم في الوقت الحالي في توليد الكهرباء وضخ المياه وطحن الحبوب، وقد تزايد الاهتمام بطاقة الرياح عالميا، وتتمتع الدول العربية بطاقة عالية من الرياح مثل الكويت والسعودية ولبنان والمغرب وتونس ومصر.
- **الطاقة الهيدروليكية أو الكهرومائية**

هي الطاقة التي يتم الحصول عليها من الأنهار وغيرها من تيارات المياه العذبة.

- **الكتلة الحيوية والغاز الحيوي** هي الطاقة المستخرجة من المواد العضوية وهي في الاساس مادة عضوية مثل المخلفات كالخشب والمخلفات الزراعية والمخلفات الحيوانية، وتقوم بتحويل طاقة الشمس المخزنة في النباتات الى طاقة متجددة .
- **الطاقة الحرارية الأرضية** هي الطاقة الحرارية من داخل الأرض، وتتكون من المياه الساخنة أو البخار الحار والتي لا زالت قيد الدراسة حتى الآن.
- **طاقة المد والجزر** هي الطاقة التي يتم الحصول عليها من المد والجزر. فهي تُستمد من ارتفاع مياه البحار والمحيطات وانخفاضها، وتعد هذه الطاقة حديثة نوعًا ما.
- **الإيثانول الحيوي** هو وقود عضوي مناسب للمركبات ويتم الحصول عليه من تخمير الغطاء النباتي.
- **وقود الديزل الحيوي** هو الوقود العضوي للمركبات، من بين تطبيقات أخرى والتي تم الحصول عليها من الزيوت النباتية .

2.3 حالة الطاقة المتجددة بالدول العربية:

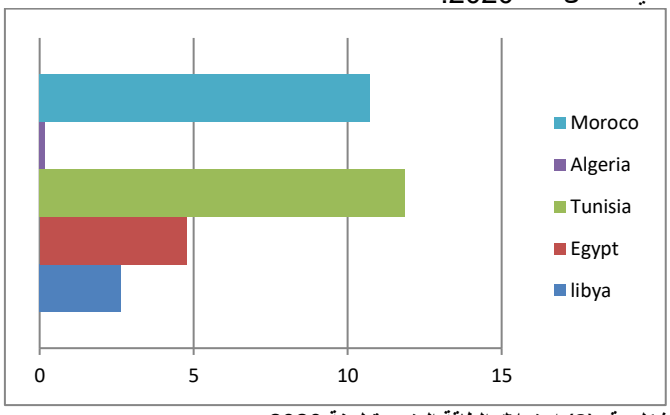
تعتمد الدول العربية على مصادر الطاقة الأحفوري لإنتاج الطاقة وعلى قدر ضئيل من الطاقة المتجددة، وقد تزايد الإهتمام بالطاقات المتجددة مؤخراً حيث قطعت الدول العربية شوطاً في استخدام الطاقات المتجددة (آسيا, الامم المتحدة ، الاسكوا ، ESCWA ،اللجنة الإقتصادية والإجتماعية لغربي 2019، 11) ويبين الشكل رقم (1) أن المغرب وتونس ومصر والجزائر قد قطعت شوطا في مجال استغلال الطاقة المتجددة ، ويرجع انخفاض استهلاك الطاقة في ليبيا والجزائر بسبب إعتمادهم على الطاقة غير المتجددة باعتبارهم أكبر دولة نفطية عربية في دول العينة، ويمثل المحور الافقي في الشكل سنوات الدراسة ويشير المحور العمودي حجم استهلاك الطاقة بدول العينة .



شكل رقم (1) يوضح استهلاك الطاقة المتجددة بالدول العربية.

*المصدر: إعداد الباحثتان اعتمادا على بيانات البنك الدولي.

ويوضح الشكل (2) حجم استهلاك الطاقة المتجددة بكل دولة عربية لسنة 2020م ، ويشير المحور الافقي لحجم استهلاك الطاقة المتجددة لدول العينة خلال سنة 2020.



شكل رقم (2) استهلاك الطاقة المتجددة لسنة 2020.

*المصدر: إعداد الباحثتان اعتمادا على بيانات البنك الدولي.

حيث يبين الشكل أن المغرب تحتل المرتبة الأولى في استهلاك الطاقة المتجددة، يليها تونس وتحتل ليبيا المرتبة ما قبل الأخيرة في استهلاك الطاقة وفقا لبيانات منظمة الطاقة العالمية (البنك الدولي، 2020) .

3.3 المؤسسات والتشريعات المنظمة للطاقات المتجددة بالدول العربية:

تعتبر تشريعات الطاقة الاساس التي تقوم عليه أسواق الطاقة عامة والمتجددة خاصة، وتنظم القوانين الاطر المؤسساتية وآليات التمويل والتنفيذ والمتابعة والتقييم، وفيما يلي عرض للقوانين والتشريعات بالدول العربية (آسيا, الامم المتحدة ، الاسكوا ، ESCWA ،اللجنة الإقتصادية والإجتماعية لغربي 2019، 16)

جدول رقم (1) القوانين والتشريعات المنظمة للطاقات المتجددة

الدول	الإطار المؤسسي	القوانين والتشريعات	التمويل
مصر	المجلس الأعلى للطاقة وهيئة الطاقة الجديدة والمتجددة	قرار المجلس الاعلى للطاقة 2011. قرار مجلس الوزراء 2014. القانون رقم 203 لسنة 2014. قرار مجلس الوزراء رقم 2598 لسنة 2016.	القطاع الخاص ومؤسسات دولية
المغرب	معهد بحوث الطاقة الشمسية. الوكالة الوطنية لتنمية الطاقات المتجددة وترشيد الطاقة. الوكالة المغربية للطاقة الشمسية	صدر قانون رقم 09.13 لسنة 2011 الخاص بالطاقات المتجددة، بغية تنمية وتكييف قطاع الطاقات المتجددة مع التطورات التكنولوجية	شركة الاستثمارات الطاقية بهدف تطوير الطاقة المتجددة بموجب قانون 08 -40 سنة 2008
تونس	تأسست الوكالة الوطنية للتحكم في الطاقة سنة 1985.	صدر القانون رقم 48 لسنة 1985 يتعلق بتشجيع البحث عن الطاقات المتجددة وفي عام 2005 تم استحداث نظام التحكم في الطاقة. وفي عام 2005 اعتمدت الحكومة نظام بروسول والذي يعمل على نشر وتطوير سوق السخانات الشمسية وثم إنشاء سوق للخلايا الشمسية.	تم انشاء وفقا لأمر حكومي رقم 983 لعام 2017 بشأن صندوق الانتقال الطاقي وصندوق الاستثمار في الطاقة
ليبيا	مركز بحوث ودراسات الطاقة الشمسية عام 1978 بغرض القيام بالأبحاث والدراسات العلمية في مجال الطاقة الشمسية وفي عام 2007 الجهاز التنفيذي للطاقات المتجددة.	لا توجد قوانين خاصة بالطاقات المتجددة.	جاري التعاون مع البنك الدولي لوضع استراتيجية متكاملة للطاقة المتجددة.

*المصدر: تقرير اللجنة الاقتصادية والاجتماعية لغربي آسيا -الامم المتحدة 2019.

4. منهجية التقدير:

1.4 النموذج القياسي:

لتقدير العلاقة بين استهلاك الطاقة والنمو الاقتصادي سوف يتم استخدام دالة الإنتاج *Cobb-Douglas* وذلك باستخدام عناصر الإنتاج المعروفة مثل رأس المال والعمل بالإضافة إلى إدراج متغير الطاقة المتجددة والطاقة غير المتجددة لدراسة أثرهما على النمو الاقتصادي، أسوة بالدراسات التطبيقية السابقة التي استخدمت هذه المنهجية.

وتأخذ دالة الإنتاج *Cobb-Douglas* الشكل التالي:

$$y = A\,K^{\propto}\,l^{\beta} \qquad (*)$$

والجدير بالذكر أن دالة الإنتاج من الأساليب المهمة في تحليل مكونات العملية الإنتاجية، حيث يساعد التطبيق القياسي لدالة الإنتاج في تقدير مؤشرات هامة وذات دلالة إحصائية تساهم في رسم الخطط اللازمة للنمو الاقتصادي والتنمية. حظيت دالة الإنتاج باستخدام كبير من الباحثين في المجال الاقتصادي لدراسة أثر التغير في عناصر العملية الإنتاجية، وتأسيساً على النظرية الكلاسيكية الحديثة فإن دالة الإنتاج تضم ثلاثة عناصر هي العمل ورأس المال والتكنولوجيا، والتي بدورها تشكل مدخلات العملية الإنتاجية، وتدمج بعضها ببعض للحصول على المخرجات وهي كمية الإنتاج *(Y)*، وحيث أن عنصر التكنولوجيا في دالة الإنتاج من الممكن أن يشير للعديد من العناصر غير المحددة والتي تدخل في العملية الإنتاجية بما يعرف بإجمالي إنتاجية عوامل الإنتاج (Rommer,2001).

حيث تكون دالة الإنتاج لغرض تحقيق هدف الدراسة على النحو التالي:

$$y = K^{\propto}\,l^{\beta}\,RE^{\delta}\,NRE^{\omega}\,e^{\mu} \quad \ldots\ldots \quad (1)$$

وبإخذ لوغاريتم طرفي المعادلة يتم تحويل المعادلة للشكل الخطي، تصبح على النحو التالي:

$$\ln y_t = c + \propto \ln k + \beta \ln l + \delta \ln RE + \omega \ln NRE + \mu \qquad \ldots\ldots\ldots \quad (2)$$

أن الشكل الخطي لدالة الإنتاج يوفر لنا نتائج متسقة، حيث تعبر كل من المعلمات المراد تقديرها عن المرونة فيما يتعلق باستهلاك الطاقة ورأس المال والعمل.

وبتوظيف البيانات المقطعية عبر الزمن لهذه الدراسة يكون النموذج القياسي المستخدم في التقدير كالتالي:

$$\ln y_{it} = c + \propto \ln k_{it} + \beta \ln l_{it} + \delta \ln RE_{it} + \omega \ln NRE_{it} + \mu_{it} \qquad \ldots\ldots\ldots \quad (3)$$

حيث تشير c إل الحد الثابت ويدل كل من الرمز i و t إلى الدولة والزمن على التوالي وهو 30 سنة (1990-2020)

Y هو الناتج المحلي الإجمالي تعبيراً على النمو الاقتصادي *(GDP)* .

K هو التكوين الرأسمالي الثابت.

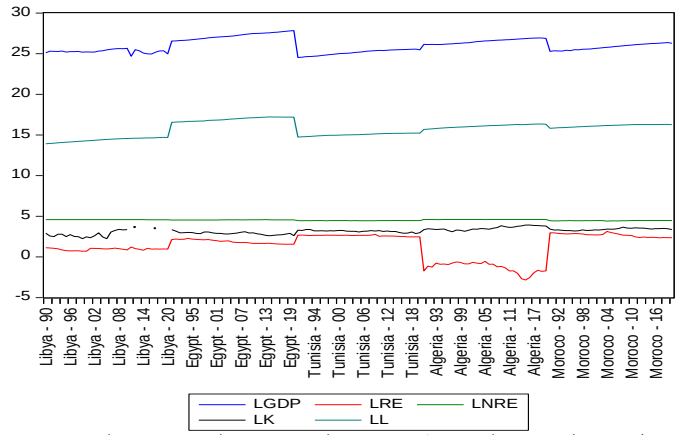
RE كمية الاستهلاك من الطاقة المتجددة.
NRE كمية الاستهلاك من الطاقة غير المتجددة.
بينما *L* تشير إلى عنصر العمل، ويُعبر عن حد الخطأ للنموذج القياسي بالرمز μ، وتشير كل من: $\propto, \beta, \delta, \omega$ إلى المعلمات المراد تقديرها والتي تمثل مرونات عناصر الإنتاج، ويدل الرمز c على الحد الثابت.

2.4 المتغيرات المستخدمة في التقدير ومصادرها:

تستخدم الدراسة المتغيرات الآتية للتعبير على عناصر الإنتاج كمدخلات للعملية الإنتاجية:

جدول رقم (2) وصف متغيرات الدراسة

المتغير	البيان	المصدر
GDP	الناتج المحلي الإجمالي، تعادل القوة الشرائية (بالدولار الدولي الثابت لعام 2017)	البنك الدولي قاعدة بيانات مؤشرات التنمية العالمية
K	اجمالي تكوين رأس المال ويتكون من النفقات على الإضافات إلى الأصول الثابتة للاقتصاد وصافي التغييرات في مستوى المخزون والأصول الثابتة، وإنشاء الطرق والسكك الحديدية ونحوها بما في ذلك المدارس والمكاتب والمستشفيات والمساكن الخاصة والمباني التجارية والصناعية.	بيانات الحسابات القومية للبنك الدولي، وبيانات الحسابات القومية لمنظمة التعاون الاقتصادي والتنمية.
L	القوى العاملة وتتألف من الأشخاص الذين تبلغ أعمارهم 15 عامًا أو أكثر والذين يقومون بتوفير العمالة لإنتاج السلع والخدمات خلال فترة محددة.	منظمة العمل الدولية، قاعدة بيانات ILOSTA. تحديث يونيو 2021.
RE	استهلاك الطاقة المتجددة (% اجمالي الاستهلاك النهائي للطاقة).	قاعدة بيانات البنك الدولي والوكالة الدولية للطاقة.
NRE	استهلاك طاقة الوقود الأحفوري (% من الإجمالي)	إحصاءات وكالة الطاقة الدولية.



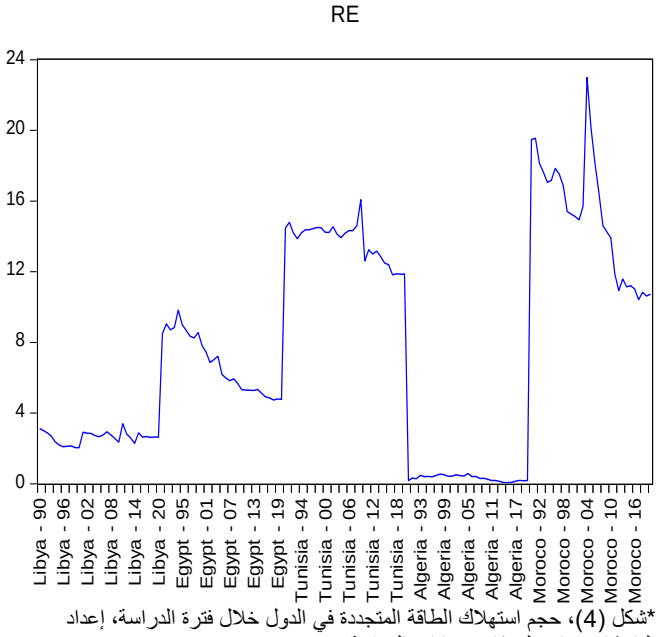
*شكل (3)، المتغيرات المستخدمة في نموذج التقدير، إعداد الباحثتان بناء على قاعدة بيانات الدراسة.

1990م - 2020م، حيث يتميز هذا النوع من البيانات بالخصائص التالية:(Gujarati, 2008)

1. تُمكننا من التحكم في عدم التجانس للمتغيرات الفردية (Hetrigenecity) وذلك بأخذه بعين الاعتبار عند التقدير.
2. توفر لنا معلومات أكثر بالإضافة للمزيد من التباين والكفاءة مع درجات حرية أكبر، وأقل تداخل (Collinearity) بين المتغيرات.
3. تتميز بجودة التحليل مقارنة باستخدام البيانات المقطعية فقط (Cross-Section)، أو استخدام بيانات السلاسل الزمنية فقط (Time-Series)، وذلك من خلال القدرة على تحديد وقياس التأثيرات التي لا يمكن تحديدها عن طريقهما.
4. بناء نموذج البيانات المقطعية عبر الزمن، يتيح إزالة تأثير أشكال معينة من تحيز المتغيرات المحذوفة في نتائج الانحدار.

أما فيما يتعلق بالمتغير الرئيسي وهو استهلاك الطاقة المتجددة من إجمالي الطاقة المستهلكة في الدول قيد الدراسة، يلاحظ تذبذب هذا المتغير خلال سنوات الدراسة في الدول الخمس، ويغلب على هذا المتغير الانخفاض في الحالة الليبية والجزائرية في إشارة لضعف استخدام المصادر الطبيعية في كلا الدولتين، حيث لم تتجاوز نسبة استهلاك الطاقة المتجددة في ليبيا 3.1%من اجمالي استهلاك الطاقة، الأمر الذي يدعو للتساؤل في بلد يحوي المقدرات اللازمة التي من شأنها المساهمة الفعالة في انتاج الطاقة النظيفة والمساهمة في إرساء دعائم الاقتصاد الأخضر. الشكل التالي يوضح حجم استهلاك الطاقة المتجددة نسبة من اجمالي استهلاك الطاقة في الدول محل الدراسة.

وتغطي البيانات الدول العربية الواقعة في الشمال الأفريقي، وهي ليبيا ومصر وتونس والجزائر والمغرب، وسيتم التركيز على الدولة الليبية عند تفسير نتائج التقدير للعلاقة بين المتغير التابع ويشار إليه في الدراسة بالناتج المحلي الإجمالي *(GDP)* والمتغيرات المستقلة.
تعتمد الدراسة على البيانات المتاحة وفقا للمصادرة المذكورة في الجدول أعلاه، وهي من البيانات المقطعية عبر الزمن غير المتوازنة Unbalanced Panel Data وذلك بسبب عدم توفر بيانات بعض سنوات الدراسة فيما يخص متغير التكوين الرأسمالي الثابت *(LK)* للدولة الليبية، الشكل التالي (3) يوضح منحنيات متغيرات الدراسة:
وإذ تتبع الدراسة في تقدير العلاقة بين المتغيرات التفسيرية والمتغير التابع منهج البيانات المقطعية عبر الزمن غير المتوازنة خلال الفترة



*شكل (4)، حجم استهلاك الطاقة المتجددة في الدول خلال فترة الدراسة، إعداد الباحثتان، بناء على قاعدة بيانات الدراسة.

3.4 الوصف الإحصائي لمتغيرات النموذج:

اعتماداً على البرنامج الإحصائي المستخدم في التقدير، يعرض الجدول (3)، الوسط والوسيط والانحراف المعياري للمتغيرات بالإضافة إلى أكبر وأصغر قيمة وعدد المشاهدات.
كما تظهر احتمالية Jargue-Bera أصغر من 0.05 مما يعني أن المتغيرات لا تتبع التوزيع الطبيعي، ويشير عدد المشاهدات لـ 155 مشاهدة لكل متغير، عدا متغير التكوين الرأسمالي الثابت وذلك بسبب عدم توفر قيم بعض السنوات لليبيا.

4.4 اختبارات السكون (Stationary Tests):

يواجه العديد من الباحثين مشكلة عدم استقراريه وسكون البيانات المستخدمة في التحليل القياسي، حيث أن الوسط الحسابي والتباين للسلاسل الزمنية تتغيران عبر الزمن، لذا يتم التعرف على درجة السكون من خلال اختبار جذر الوحدة (Unit Root Test) لهذه البيانات وتحديد درجة التكامل وذلك تجنباً لنتائج تقدير زائفة. ويعتبر اختبار التكامل بين البيانات المقطعية عبر الزمن من أهم مراحل

جدول (3)، الوصف الإحصائي

	GDP	RE	NRE	K	L
Mean	2.79E+11	7.709082	93.61583	25.65611	9761124.
Median	1.61E+11	6.188900	95.08310	25.01377	8669900.
Maximum	1.22E+12	23.00070	99.97792	50.78072	29972669
Minimum	4.48E+10	0.059000	81.49045	9.481002	1110018.
Std. Dev.	2.63E+11	6.142881	5.629888	8.837849	8021133.
Skewness	1.702207	0.310036	−0.262846	0.628536	1.095424
Kurtosis	5.411274	1.767487	1.403036	3.354185	3.376652
Jarque−Bera	112.4026	12.29394	18.25541	10.44728	31.91504
Probability	0.000000	0.002140	0.000109	0.005388	0.000000
Sum	4.33E+13	1194.908	14510.45	3771.448	1.51E+09
Sum Sq. Dev.	1.06E+25	5811.188	4881.129	11403.71	9.91E+15
Observations	155	155	155	147	155

*وصف المتغيرات، إعداد الباحثتان بناء على مخرجات EVIEWS 9.

التحليل للنماذج القياسية، إذ يعتبر إضافة البعد الفردي (الدول) إلى البعد الزمني (السنوات) ذا أهمية قصوى في تحليل السلاسل الزمنية.
الجدول رقم (4) يشير إلى نتائج اختبار جذر الوحدة للمتغيرات المستخدمة في النموذج القياسي، حيث تجمع الاختبارات التي تم إجراءها بين فرضيتي جذر الوحدة وانعدام الأثار الفردية، وتأسيساً على ذلك تعتبر المتغيرات ساكنة إذا كانت P-Value <0.05 حيث يتم رفض الفرضية الصفرية التي تقضي بعدم سكون المتغير على المستوى وقبول الفرضية البديلة التي تفيد بعكس ذلك، وعندما تكون المتغيرات غير ساكنة في المستوى نقوم بإعادة اجراء اختبار السكون عند الفرق الأول.
يتضح من الجدول التالي أن كلٌ من *LGDP, LL* هما مستقران على المستوى *I(0)*، بينما باقي المتغيرات المستخدمة في نموذج التقدير *LK, LRE, LNRE* متغيرات مستقرة على الفرق الأول *I(1)*.

جدول (4)، اختبارات جذر الوحدة

	Common Unit Root Test		*Individual Unit Root test*					
Variables	*Levin, Lin & Chu*	*(P)*	*Im, Pesaran & Shin*	*(P)*	*ADF Fisher Chi−Square*	*(P)*	*PP Fisher Chi−Square*	*(P)*
				Level				
LGDP	*−2.6566*	*0.0039*	*−1.0448*	*0.1481*	*19.2028*	*0.0378*	*18.2861*	*0.050*
LK	*−0.8192*	*0.2063*	*0.0885*	*0.5353*	*8.6321*	*0.5673*	*9.4227*	*0.4925*
LL	*−4.3202*	*0.0000*	*−2.2232*	*0.0131*	*23.2317*	*0.0099*	*62.6923*	*0.0000*
LRE	*−2.5162*	*0.0059*	*−0.4532*	*0.3252*	*12.2374*	*0.2637*	*8.6157*	*0.5689*
LNRE	*−1.2533*	*0.1050*	*−1.0835*	*0.1393.*	*14.0940*	*0.1687*	*14.4222*	*0.1546*
				First Diff.				
LK	*−9.8128*	*0.0000*	*−9.6692*	*0.0000*	*88.1091*	*0.0000*	*100.803*	*0.0000*
LRE	*−11.0219*	*0.0000*	*−9.7705*	*0.0000*	*91.3631*	*0.0000*	*133.788*	*0.0000*
NRE	*−13.6227*	*0.0000*	*−12.5548*	*0.0000*	*117.311*	*0.0000*	*124.888*	*0.0000*

إعداد الباحثتان بناء على مخرجات البرنامج الإحصائي (EVIEWS 9).

5.4 اختبار التكامل المشترك بين متغيرات النموذج (Co-integration test):

الغرض من هذا الاختبار معرفة وجود التكامل المشترك بين متغيرات النموذج القياسي، بهدف التنبؤ بطبيعة العلاقة بين تلك المتغيرات في الأجل الطويل، ولغرض التحقق من هذه العلاقة تم استخدام اختبار Augmented Dicky-Fuller، وفي حالة كانت P-Value<0.05 تقبل الفرضية البديلة التي تفيد بوجود علاقة تكامل مشترك طويلة الأجل بين المتغيرات. الجدول التالي يبن نتيجة هذا الاختبار:

جدول (5)، نتيجة اختبار التكامل المشترك

ADF	*t-statistic*	*Prob.*
	-2.197764	*0.0140*

*إعداد الباحثتان بناء على مخرجات البرنامج الإحصائي (EVIEWS 9).

حيث أن متغيرات الدراسة مرتبطة بعلاقة توازنيه طويلة الأجل، فإن النموذج المناسب للتقدير هو نموذج الانحدار الذاتي لفترات الإبطاء الموزعة Autoregressiv Distributed Lag Model للبيانات الزمنية المقطعية، وذلك لقياس ديناميكية العلاقة وتحديد الآثار طويلة وقصيرة الأجل بين متغيرات النموذج المستخدمة في الدراسة.

6.4 منهجية الانحدار الذاتي لفترات الإبطاء الموزعة (ARDL):

تم تطوير هذه المنهجية من قبل Pesaran, Shin & Smith في سنة 2001م والتي تسمح بالكشف عن وجود علاقة التكامل المشترك بين المتغيرات المستخدمة في النموذج القياسي، ويتم تطبيق هذه المنهجية في حالة تكون فيها المتغيرات مختلفة درجات ا لتكامل، ويُعد هذا الأسلوب من الأساليب الديناميكية التي تسمح للمتغيرات المستقلة بأن تكون متغيرات تابعة تتحدد من داخل النموذج (Pesaran. Et al, 2001).
وبعد أن توصلنا إلى وجود علاقة تكاملية طويلة الأجل بين متغيرات النموذج، تم اختيار أنسب فترات الإبطاء، وكانت فترة الإبطاء المثلى واحدة (1, 1, 1, 1) في النموذج المعد لتقدير العلاقة بين الناتج المحلي الإجمالي وعناصر العلمية الإنتاجية وفقاً لدالة الإنتاج.

تم تقدير العلاقة قصيرة الأجل وطويلة الأجل بمنهجية ARDL، بهدف التوصل إلى النتائج التي تمكننا من تفسير العلاقة بين النمو الاقتصادي وكمية الطاقة المستهلكة من المصادر الطبيعة والمصادر غير الطبيعية بالإضافة إلى عنصري رأس المال والعمل، **وقد جاءت نتائج التقدير لدالة Cobb-Douglas على النحو التالي:**

يظهر الجدول رقم (6) أن احتمالية معلّمات النموذج في الأجل الطويل جاءت جميعها معنوية عند مستوى أقل من 10% ، الأمر الذي يؤيد الفرضية التي قامت عليها الدراسة بأن هناك علاقة ذات دلالة إحصائية بين النمو الاقتصادي واستهلاك الطاقة المتجددة في دول العينة، وجود تأثير موجب ومعنوي لاستهلاك الطاقة على النمو الاقتصادي، بسبب أن الطاقة تعد مدخل أساسي للعملية الإنتاجية، مما يجعل من استهلاك الطاقة تأثيراً موجبا على النمو الاقتصادي، ونظرياً تدعم هذه النتيجة نموذج (Solow, 1974) الذي طور دالة الإنتاج بإدخال متغير الطاقة كأحد مدخلات العملية الإنتاجية، وتطبيقياً تتفق هذه النتيجة مع نتائج عدد من الدراسات السابقة. إذ تشير نتائج التقدير أن مرونة الناتج المحلي الإجمالي بالنسبة لاستهلاك الطاقة المتجددة نحو 0.27 أي أن زيادة استهلاك الطاقة المتجددة بنسبة 100% يؤدي إلى زيادة النمو الاقتصادي بنسبة 27%، كما بلغت مرونة استهلاك الطاقة من الوقود الأحفوري - 7.49 بمعنى أن أي زيادة في استهلاك الطاقة غير المتجددة بـ 100% يؤدي إلى انخفاض في النمو الاقتصادي بنسبة 7.49% على المدى الطويل بسبب استهلاك واستنفاد مصدر من مصادر الدخل القومي في دول العينة. ويرتبط الناتج المحلي الإجمالي بعلاقة طردية موجبة مع كلٌ من عنصر العمل وعنصر رأس المال حيث أن زيادة التكوين الرأسمالي والعمل بنسبة 100% تؤدي إلى زيادة الناج المحلي الإجمالي بـ 0.41%، 2.46% على التوالي، وهو ما يتفق مع النظريات الاقتصادية ونتائج الدراسات السابقة في ذات المجال، حيث لا مجال للشك في قدرة العنصر البشري والمادي في تحفيز والرفع من الناتج المحلي الإجمالي لدول العينة.

أما عن حد تصحيح الخطأ (Error Correction Model) ECM(-1) والذي يقيس سرعة التعديل إلى التوازن في الأجل الطويل، فقد كانت المعلمة (0.0867-) ، حيث أن العلامة السالبة تزيد من صحة ودقة العلاقة التوازنية في الأجل الطويل، مما يعني أن آلية تصحيح الخطأ موجودة وذات معنوية عالية (أقل من 0.05)، وتدل على أن 8.67% من أخطاء المدى القصير يتم تصحيحها في الأجل الطويل حتى تعود إلى التوازن.

جدول (6)، نتائج التقدير النموذج القياسي

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
Long Run Equation				
LRE	0.276072	0.159471	1.731169	0.0862
LNRE	-7.490364	2.771182	-2.702949	0.0079
LK	0.416875	0.227096	1.835677	0.0690
LL	2.461673	0.229275	10.73675	0.0000
Short Run Equation				
COINTEQ01	-0.086763	0.029438	-2.947345	0.0039
D(LRE)	-0.078000	0.011754	-6.635882	0.0000
D(LNRE)	-4.855176	4.566377	-1.063245	0.2899
D(LK)	0.019269	0.011260	1.711239	0.0898
D(LL)	0.794301	0.622857	1.275255	0.2048
C	1.707597	0.578700	2.950746	0.0039

*مخرجات Eviews 9

وفي الجانب الآخر لا تتفق إشارة المعلمة ومستوى المعنوية لمتغير استهلاك الطاقة المتجددة في الأجل القصير مع مثيلتها في الأجل الطويل، حيث جاءت الإشارة سالبة مع دلالة معنوية عالية أصغر من0.05، وذلك أن استهلاك الطاقة المتجددة لا يؤثر بشكل إيجابي على النمو الاقتصادي في الأجل القصير، ولكن يظهر أثره جلياً في المدى الطويل على النمو الاقتصادي لدول العينة. وبالتركيز على حالة الدولة الليبية وبالنظر إلى المعلمات المقدّرة للمتغيرات المستقلة في الأجل القصير من خلال مخرجات Eviews 9، يتضح من الجدول (7) لـ:

(Cross Section Short Run Coefficient) أن العلاقة بين النمو الاقتصادي و استهلاك الطاقة المتجددة في ليبيا علاقة سلبية ذات معنوية عالية أقل من 0.05 %، الأمر الذي يشير إلى انخفاض استهلاك الطاقة المتجددة بحيث لن تستطيع هذه الكمية تحفيز النمو الاقتصادي في الأجل القصير، وبالنسبة للطاقة غير المتجددة فأنها ارتبطت بعلاقة سلبية مع النمو الاقتصادي الليبي في الأجل القصير لفترة مبطأة واحدة حيث تشير قيم المعلمة المقدرة إلى أن زيادة استهلاك الطاقة بنسبة 100% يترتب عليه انخفاض النمو الاقتصادي بحوالي 23 % ، ويعود ذلك إلى دعم استهلاك الطاقة في ليبيا خلال فترة الدراسة والذي ينتج عنه خسائر يتكبدها النمو الاقتصادي، بالإضافة إلى التكاليف المالية المباشرة للدعم المالي الموجه إلى قطاع الطاقة، يؤدي إلى زيادة الاستهلاك وانخفاض الحافز للحفاظ على الطاقة وترشيد استهلاكها، بينما تبين النتائج العلاقة الإيجابية بين النمو الاقتصادي وعنصر التكوين الرأسمالي الثابت لأهمية تكوين رأس المال الثابت في دعم الاستثمار واستغلال الموارد وتوظيفها في رفع مستوى معيشة الأفراد بالمجتمع وانخفاض البطالة الذي يعود إيجاباً على الاقتصاد الليبي .

جدول (7)، تقدير النموذج في الأجل القصير لـ ليبيا

Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.062535	0.002256	-27.71815	0.0001
D(LRE)	-0.097356	0.024683	-3.944295	0.0291
D(LNRE)	-23.02034	1440.701	-0.015979	0.9883
D(LK)	0.010625	0.002612	4.067250	0.0268
D(LL)	3.228945	13.94331	0.231577	0.8318
C	1.373799	1.631437	0.842079	0.4616

*مخرجات Eviews 9

5.التوصـيات:

بناء على نتائج التقدير تقدم الدراسة عدداً من التوصيات التي ترى الباحثتان أن الأخذ بها من شأنه دعم قطاع الطاقة المتجددة مستقبلاً في ليبيا ودول العينة على حد سواء:

1. التوسع في تكنولوجيا الطاقة المتجددة عن طريق تشجيع الشركات للاستثمار في البحث والتطوير في مجال الطاقة المتجددة، مع معالجة الهدر في إستخدام الطاقة والعمل على ترشيد إستهلاك الطاقة في المجتمع المحلي .
2. الاستفادة من خبرات بعض الدول العربية عن طريق التعاون مع مراكز البحث والتطوير بتلك الدول.
3. تشجيع الصناعات المختلفة على استخدام التقنيات والوسائل التي تعتمد على الطاقات المتجددة من أجل منع التلوث وتحقيق تنمية مستدامة.
4. على الدولة توفير الدعم المالي اللازم لقطاع الطاقة المتجددة، وذلك بتخصيص جزء من عائدات النفط في تمويل المشاريع التي من شأنها إحداث نقلة نوعية في هذا المجال.
5. الاستغلال الأمثل لمناخ وموقع ليبيا الجغرافي في انتاج الطاقة النظيفة، والاستفادة من الدول الرائدة في مجال الطاقة المتجددة.
6. العمل على تنويع الطاقة التي يتم استخدامها لصالح احلال الطاقة المتجددة محل المصادر غير المتجددة على المدى الطويل، تشجيع وتطوير المراكز البحثية فيما يتعلق بإمكانية التوسع في استخدام الطاقة المتجددة وسبل تعزيزها في مختلف مناطق البلاد.
7. للدراسات المستقبلية عن الطاقة المتجددة، تقترح الباحثتان البحث في تحديد مصادر الطاقة المتجددة التي تساهم بشكل أكبر في النمو الاقتصادي، وبالتالي تساعد في تحديد المصدر الأكثر فائدة اقتصاديًا لتطوير هذا المصدر المتجدد.

6. الخلاصة :

تبحث هذه الدراسة عن علاقة الطاقة المتجددة بالنمو الاقتصادي قياسيا في الدول العربية (ليبيا – مصر – تونس – الجزائر – المغرب) خلال الفترة من 1990 م – 2020 م وفقا للبيانات المقطعية عبر الزمن غير المتوازنة وبتوظيف دالة الانتاج (Cobb Douglas) وتطبيق منهجية (ARDL) الانحدار الذاتي لفترات الابطاء الموزعة ، وقد توصلت الدراسة إلى وجود علاقة ذات دلالة إحصائية بين النمو

الإقتصادي وإستهلاك الطاقة المتجددة في دول العينة وذلك عند مستوى 10% ، بمعنى أن زيادة إستهلاك الطاقة المتجددة في العمليات الانتاجية بنسبة 100% يؤدي إلى زيادة النمو الإقتصادي بنسبة 27% .

7.المراجع:

- آسيا ,الامم المتحدة ، الاسكوا ، ESCWA،اللجنة الإقتصادية والإجتماعية لغربي .الطاقة المتجددة -التشريعات والسياسات في المنطقة العربية .بيروت :الأمم المتحدة، 2019 .
- البنك الدولي، 2018 .
- الحسنين، مروة" .تحليل العلاقة بين استهلاك الطاقة المتجددة والنمو الاقتصادي في المغرب العربي باستخدام نموذج الانحدار الذاتي للابطاء الموزع غير الخطي ".NARDLدراسات، ابريل 2021.
- بوعتلي، محمد، دراسة قياسية لتأثير استهلاك الطاقات المتجددةعلى النمو الاقتصادي في دول المغرب العربي 2019 .
- حسن ,ضياء محمد، أثر استهلاك الطاقة في النموالاقتصادي في مصر .مجلة البحوث التجارية -كلية التحارة جامعة الزقازيق، 2020.
- راتول، محمد و محمد مداحي، صناعة الطاقات المتجددة بألمانيا وتوجه الجزائر لمشاريع الطاقة المتجددة لتأمين إمدادات الطاقة الأحفورية وحماية البيئة "حالة مشروع ديزرتاك ، جامعة قاصدي مرباح، 140 .2012
- زهير ,عماري" .مصادر النمو الاقتصادي لدول العالم العربي بين إشكالية كفاءة استغلال الموارد والتكاليف البيئية دراسة مقارنة الجزائر قطر ".2012-1990الجمعية الدوليةللاقتصاد الإسلامي، 2015 .
- صندوق النقد العربي ، تقرير افاق الاقتصاد العربي،2020.
- المركز الاقليمي للطاقة المتجددة دليل الطاقة المتجددة وكفاءة الطاقة في الدول العربيةRCREEE |
- عاشور، عبدالله، انتاج واستهلاك الطاقة في الاقتصاد الليبي- دراسة تحليلية من منظور التنمية المستدامة، رسالة ماجستير غير منشورة، جامعة بنغازي، 2012.

- Favara, Giovanni. "An Empirical Reassessment of the Relationship between Finance and Growth∗." 2006.
- Fu, Q., Álvarez-Otero, S., Sial, M. S., Comite, U., Zheng, P., Samad, S., & Oláh, J. (2021). Impact of renewable energy on economic growth and CO2 emissions—evidence from BRICS countries. *Processes*, *9*(8), 1281.
- Gujarati, D. (2008). N, 2003**, Basic Econometrics**, New York: MeGraw-Hill, 363-369.
- international renewable energy Agency. "ENERGY PROFILE." 2020.
- Lin hung, Pin, (2014) Renewable Energy Consumption and Economic Growth in Nine OECD Countries: Bounds Test Approach and Causality Analysis**,** volume 2014, Scientific World Journal.
- Pesaran, Hashem; Yongcheol Shin & Richard J. Smith (2001). **Bounds Testing Approaches to the Analysis of Level Relationships**, *Journal of Applied Econometrics*, 16 (3): 289-326.
- Rommer, David (2001). ***Advanced Macroeconomics***, 2nd Edition McGraw-Hill Companies, Inc.
- Solow, R. M. (1974). **The economics of resources or the resources of economics**. In *Classic papers in natural resource economics* (pp. 257-276). Palgrave Macmillan, London.

.

List of Authors

قائمة المؤلفين

Name (Organization), Paper ID