

Thermoeconomic Optimisation of Steam Condenser for Combined Cycle Power Plant

Muammer Alus
Faculty of Engineering
Mechanical and Industrial
Engineering Department
University of Elmergib
Alkhoms, Libya

Mohamed Elrawemi
Faculty of Engineering
Mechanical and Industrial
Engineering Department
University of Elmergib
Alkhoms, Libya

Faraj Eldabee
Faculty of Engineering,
Mechanical and Industrial
Engineering Department
University of Elmergib
Alkhoms, Libya

Abstract: In this study, the effect of condenser pressure on condenser area and the thermoeconomic parameters is investigated. The objective function of the thermoeconomic optimisation is to minimise the production costs per unit of electrical output and to maximise the annual cash flow of the power plant. Therefore, the thermoeconomic parameters are generated for different range of condenser pressure. The case 283 MW combined cycle gas turbine (CCGT) power plant with typical design of commercial combined cycle power plant was demonstrated as an example. The optimised combined cycle was compared with a regular CCGT plant. The results indicated that the thermoeconomic optimisation procedure used in this study led to a significant improvement for the economic parameters, where the annual cash flow was increased by 1.1 M\$, and the production cost of electricity was decreased by 0.043 cent-dollar per kilowatt-hour.

Keywords: thermoeconomic; optimisation; steam; condenser; combined cycle gas turbine (CCGT)

1. INTRODUCTION

The condenser is a heat transfer device or unit used to condense a substance from its gaseous state to its liquid state, typically by cooling it. The main use of the condenser is to receive the exhaust steam from steam turbine and condense it. The benefit of this process is the energy, which can be utilised instead of exhausting it to the atmosphere. The steam condenser generally condenses the steam to a pressure below atmospheric. This allows the turbine to generate more work. The condenser also converts discharge steam back to feed water, which is returned to steam generators. In the condenser, the latent heat of condensation is conducted to the medium flowing through the cooling tubes. The thermal power plants are designed based on required conditions, however actually the inlet conditions are not as per the designed conditions.

In practical situations, once the power plants are installed there are a lot of constrains. This tends to reduce or increase the output power and heat rate of thermal power plants. As a result of these conditions, the designed power and heat rate are never achieved [1].

Moreover, the designed features of the condenser have a significant impact, but it is expensive to replace them when the plant is operating. At present, it is possible to improve the design for new plant due to technology development. The considerations that the designers should take into account are the weather condition (cooling water temperature) and the economic condition (fuel price), where the power plant is installed.

Different approaches can be found in the literature regarding the performance of steam condenser. The impact of cooling water temperature on the efficiency of the power plant was stated using simulator based on IAPWS-IF97. It is found that increasing the pressure in the condenser to 1 kPa decreases the efficiency between 1.0 to 1.5% [3]. Also, the parameters, which affect the performance of the condenser and the power plant, were studied. The authors found that the performance of

the condenser was varied due to the cooling water inlet temperature, flow rate of cooling water and pressure in energy [4]. Furthermore, an optimisation of a condenser in a thermal power plant was performed. The effects of the condenser pressure on the condenser area and the initial cost of the condenser was investigated. It is concluded that the area of condenser increases when the pressure of condenser decreases, accordingly the initial cost of the condenser increases [5].

The earlier studies revealed that the condenser pressure has an influence on the energy efficiency of the thermal power plants and on the area of the condenser. These aspects have prompted the authors of this research paper to focus on studying the effect of the condenser pressure on the thermoeconomic parameters.

2. THERMODYNAMIC OPTIMISATION MODEL

The objective of thermodynamic optimisation is to enhance the efficiency of the combined cycle power plant (CCGT) and to maximise the electrical power in the steam cycle (steam turbine gross power). Here, a CCGT cycle with triple-pressure HRSG will be considered. Figure 1 shows a schematic diagram of a CCGT. The compute code to calculate the heat balance of triple-pressure HRSG of CCGT was developed using FORTRAN 90. The water-steam properties were derived from the standard IAPWS-IF97. The properties of the gas turbine exhaust gases, which are combustion products of the specified fuel, were calculated according to [6]. The assumptions and the studied parameters, which are selected for the thermodynamic analysis of the plant, are presented in Table 1.

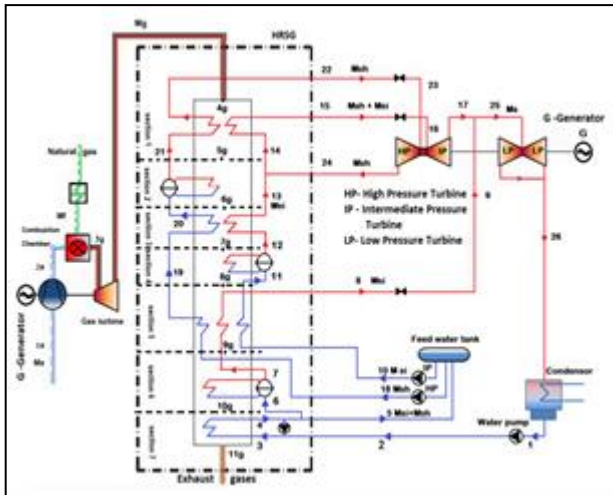


Figure 1. Gas turbine and steam turbine combined cycle – heat balance diagram [1]

Table 1- Gas turbine parameters for CCGT with the triple-pressure HRSG [1]

Parameter	Value
1. Gas Turbine Cycle (Alstom GT24/1994)	
Ambient air pressure [bar] & ambient air temperature [oC]	1.013 & 15
Electrical power at the generator output [MW]	187.7
Exhaust gas mass flow [kg/s]	445
Exhaust gas temperature at the gas turbine outlet [oC]	612
The gas turbine efficiency [%]	36.9
Lower heat value of the fuel [kJ/kg]	47141
2. Steam Turbine Cycle	
The pinch point temperature difference for HP, IP and LP [oC]	13
The minimum temperature difference between the gas turbine exhaust gases and live/reheat steam [°C]	25
Live steam pressure (HP) [bar]	104
Live steam temperature at the inlet of the HP steam turbine [oC]	545
Pressure of reheat steam (IP steam turbine) [bar]	36
Temperature of the reheat steam (IP steam turbine) [oC]	545
Pressure of the inlet LP steam turbine [bar]	5
Temperature of the superheated steam at 8 [oC]	235
Temperature of the superheated steam at 13 [oC]	320
Steam mass flow [kg/s]	73
Low-pressure steam turbine outlet (condenser pressure) [bar]	0.057
Feed water temperature at 3 [oC]	60
3. Combined Cycle Power Plant	
Net total electrical power (MW)	283.07
Net plant efficiency (%)	55.65
4. Assumption	
The isentropic efficiency of all three steam turbine parts [%]	90
The isentropic efficiencies of water pumps [%]	82
The mechanical efficiency [%]	99.5
The generator efficiency [%]	98
The heat recovery steam generator efficiency	99.3
The pressure drops for water in the economisers [%]	25

The pressure drops for steam in the reheat and super-heater tubes [%]	8
The overall heat transfer coefficients for sections of the HRSG [W/m ² K]	42.6
• economiser	43.7
• evaporator	50.0
• super-heater and reheat	
The overall heat transfer coefficient for the condenser [Wm ⁻² K ⁻¹]	2500
The inlet cooling water temperature in condenser (°C)	20

2.1 The Condenser Model

In power plants, the most commonly used condenser is the steam surface condenser. Consequently, steam surface condenser with open cycle cooling water supply system in this study is assumed.

2.1.1 Element of Surface Condenser

The heat mechanism is the condensation of saturated steam outside the tubes and heating of the circulating water inside the tubes. Figure 2 presents the heat transfer diagram of the condenser. The steam temperature (t_s) is the saturation temperature at the condenser pressure. The difference between the steam temperature (t_s) and the water inlet temperature (tw_1) is defined as the initial temperature difference (ITD). The difference between the steam temperature (t_s) and the water outlet temperature (tw_2) is known as the terminal temperature difference (TTD). The circulating water inlet temperature should be sufficiently lower than the steam saturation temperature to result reasonable values of TTD.

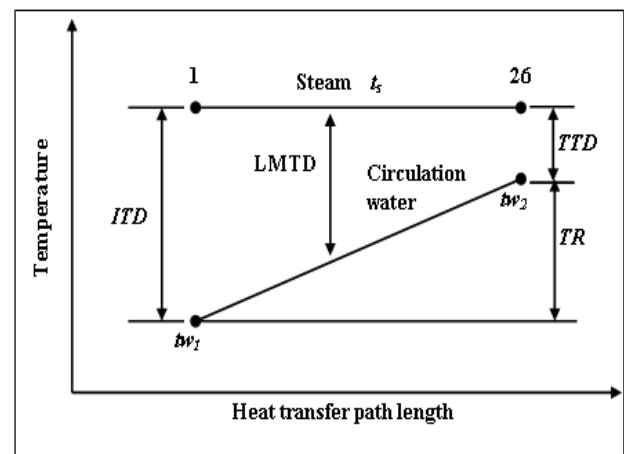


Figure 2. Heat transfer diagram of condenser

2.1.2 Calculation of Condenser Heat Transfer Area

The heat transfer rate between the cooling water and the steam vapour is the key parameter of the thermal analysis of the steam condenser. The heat from the exit steam is:

$$\dot{Q}_{ST} = \dot{m}_{ST} (h_{26} - h_1) \quad (1)$$

The heat that cooling water gets:

$$\dot{Q}_w = \dot{m}_w c_{p_w} (tw_2 - tw_1) \quad (2)$$

where \dot{m}_w represents the cooling water flow rate in (kg/s), and c_{p_w} the specific heat of the water in (kJ/kg.K).

The heat transferring in condenser is:

$$\dot{Q}_{Cond.} = A_{Cond.} \cdot U_{Cond.} \cdot LMTD_{Cond.} \quad (3)$$

where, $\dot{Q}_{Cond.}$ is the heat transferred and $U_{Cond.}$ is the heat transfer coefficient for the condenser in (W/m². K). In condenser calculation, the heat transfer coefficient $U_{Cond.}$ is assumed. LMTD Cond is the logarithmic mean temperature difference for the condenser, which is defined as:

$$LMTD_{Cond.} = \frac{TR}{\ln \frac{1}{1 - \frac{TR}{ITD}}} \quad (4)$$

where TR is the temperature rise in the condenser and defined as $(t_{w2} - t_{w1})$ and ITD is the initial temperature difference.

The condenser heat transfer area $A_{Cond.}$ was calculated according to [7]:

$$A_{Cond.} = \frac{\dot{Q}_{Cond.}}{U_{Cond.} \cdot LMTD_{Cond.}} = \frac{\dot{m}_{ST} (h_c - h_i)}{U_{Cond.} \cdot LMTD_{Cond.}} = \frac{m_w \cdot c_{p_w} (t_{w2} - t_{w1})}{U_{Cond.} \cdot LMTD_{Cond.}} \quad (5)$$

2.2 Combined Cycle Efficiency

The overall cycle efficiency can be obtained from the steam power and gas turbine powers, which also are dependent on many other variables, such as condenser pressure, steam pressures and ambient temperature. The thermal efficiency in a combined cycle plant is defined by the relation:

$$\eta_{CCGT} = \frac{W_{CCGT}}{Q_{add}} = \frac{W_{ST} (\dot{m}_{ST}, h_{cond}, \eta_{ST}) + W_{GT}}{\dot{m}_f \cdot H_i} \quad (6)$$

where \dot{m}_f is mass flow rate of fuel in (kg/s), H_i Lower heat value of the fuel (kJ/kg).

3. THERMOECONOMIC OPTIMISATION MODEL

To perform the thermoeconomic optimisation, some economic parameters should be assumed, and relations between the thermodynamic parameters and capital cost have to be established and objective functions defined.

3.1 Main Thermoeconomic Characteristics of CCGTs

The thermoeconomic optimisation was performed under the following conditions, which were taken according to [1]:

- The average life of the combined cycle power plant is 20 years.
- The power plant is operated for 7500 h a year.
- The unit selling price of generated electricity is 0.114 \$/kWh.
- Price of natural gas 0.0467 \$/kWh.
- The installed costs of the economiser, evaporator, superheater and reheat sections of the HRSG are 45.7, 34.8 and 96.2\$/m², respectively.

3.2 Functions of Component Capital Costs

The basic problem in the analysis of the economic effectiveness of investments in energy systems is the determination of capital costs. For this study, the cost functions for the major components of a combined cycle power plants were taken from literature: cost of gas turbine C_{GT} [8], cost of HRSG C_{HRSG} [9], cost of steam turbine C_{ST} , cost of pump C_{Pump} and cost of generator

C_{Gen} [10]. Cost of condenser $C_{Cond.}$ [11] is function in condenser area and is expressed as:

$$C_{Cond} = 280 \cdot A_{Cond}^{1.01} \quad (7)$$

The cost functions give net capital costs of the components. However, the net capital cost does not include: transport and assembly costs, supervising, accessories, engineering and project management, commissioning and other connected costs. Therefore, the additional correction factor R was introduced in Eq. 8 to cover these additional costs and increase in price in recent period. The R value is obtained by analysis and consulting the market. In the calculations, the value of R is assumed to be 3.0 [1].

The total capital costs (investment costs) of a combined cycle power plants are given by:

$$C_{CCGT} = R \cdot \sum_i C = R \cdot (C_{GT} + C_{HRSG} + C_{ST} + C_{Cond.} + C_{Pump} + C_{Gen.}) \quad (8)$$

The cost function of HRSG, steam turbine, condenser, pump and generator in Eq. 8 is expressed as function of operating parameters ($p_{Live\ Steam}$, $t_{Live\ Steam}$, $P_{Cond.}$, etc). Therefore, any change in these variables will cause a change in the total capital costs.

3.3 Objective Function

The annual cash flow (B) and the production cost of electricity (C_{kWh}) have been chosen as an objective function of the thermoeconomic optimisation.

The annual cash flow (B) is the difference between the annual total income (I_{tot}) and the total cost per year (C_{tot}) as pointed out by [12], [13]:

$$B = I_{tot} - C_{tot} \quad (9)$$

The production cost of electricity was defined as pointed out by [12], [13]:

$$C_{kWh} = \frac{C_{tot}}{W_{CCGT} \cdot h} \quad (10)$$

Where C_{tot} the total cost per year and the product ($W_{CCGT} \cdot h$) is the total annual energy production.

The annual cash flow (B) is the difference between the annual total income (I_{tot}) and the total cost per year (C_{tot}) as pointed out by [12], [13]:

$$B = I_{tot} - C_{tot} \quad (11)$$

The annual total income can be calculated using

$$I_{tot} = S \cdot W_{CCGT} \cdot h \quad (12)$$

where S is the selling price per unit of electricity.

The total cost per year includes the total annual fuel cost, the amortisation cost and the operating and maintenance cost, as shown in the following equation:

$$C_{tot} = C_{TF} + C_a + C_{o\&m} \quad (13)$$

The total annual fuel cost (C_{TF}) is defined as follows:

$$C_{TF} = c_f \cdot \left(\frac{W_{CCGT}}{\eta_{CCGT}} \right) \cdot h \quad (14)$$

The amortisation cost (C_a) is summarised as pointed out by [13]:

$$C_a = \frac{C_{CCGT}}{N} \quad (15)$$

The operating and maintenance cost ($C_{o\&m}$) is assumed to be 10 % of the total plant cost as pointed out by [13]:

$$C_{o\&m} = 0.10 \cdot (C_{tot}) \quad (16)$$

4. RESULTS AND DISCUSSION

The condenser pressure ($P_{Cond.}$) was in the range of 0.04 to 0.08 bar by adding 0.0005 bar incrementally in each step. The thermodynamic and the thermo-economic parameters were calculated for each step.

Figure 3 shows the combined cycle gross power as a function of condenser pressure ($P_{Cond.}$) of triple-pressure HRSG CCGTs. The result shows that the combined cycle gross power decreases by increasing $P_{Cond.}$. Decreasing the $P_{Cond.}$ will lead to higher power output for the same mass flow rate and fuel input into gas turbine unit, resulting in higher work output of the steam turbine.

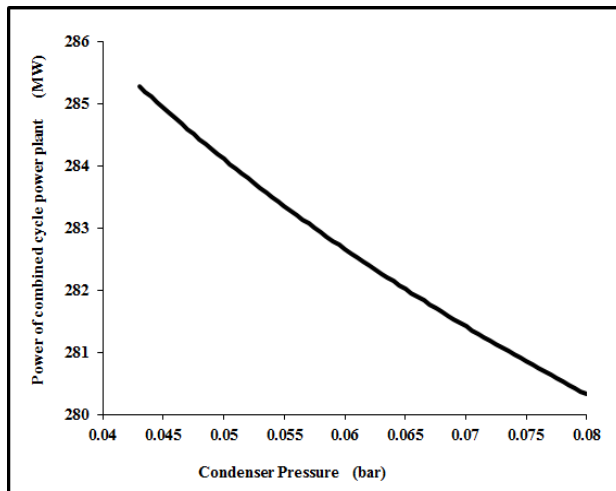


Figure 3. Effect of condenser pressure variations on the combined cycle gross power

Figure 4 shows the cost of condenser ($C_{Cond.}$) as a function on the condenser pressure ($P_{Cond.}$). Through the figure 4, the result shows that a decrease in the $P_{Cond.}$ will significantly increase the necessary heat transfer surface area of condenser ($A_{Cond.}$) and incase the initial cost of condenser ($C_{Cond.}$). On the other hand, an increase in the area of the condenser means more the heat recovered and an improvement in the CCGT thermal efficiency.

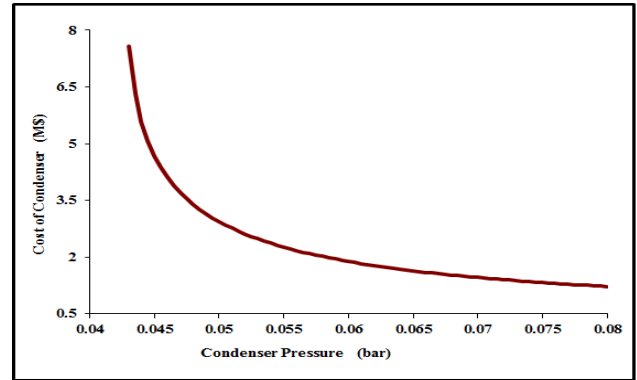


Figure 4. Effect of condenser pressure variations on the cost of condenser

Figure 5 demonstrates the effect of the condenser pressure ($P_{Cond.}$) on the annual cash flow (B), which is defined by Eq. 9. As can be seen, annual cash flow (B) increases with increase in the $P_{Cond.}$, until it reaches the optimal value and then decreases with further increase in the $P_{Cond.}$. The result shows that the optimal value for the $P_{Cond.}$, at which the maximum annual cash flow (B) is achieved, is 0.0445 bar.

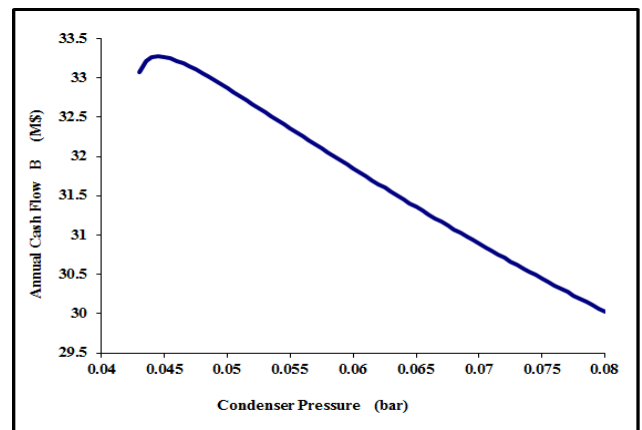


Figure 5. Effect of condenser pressure variations on the annual cash flow

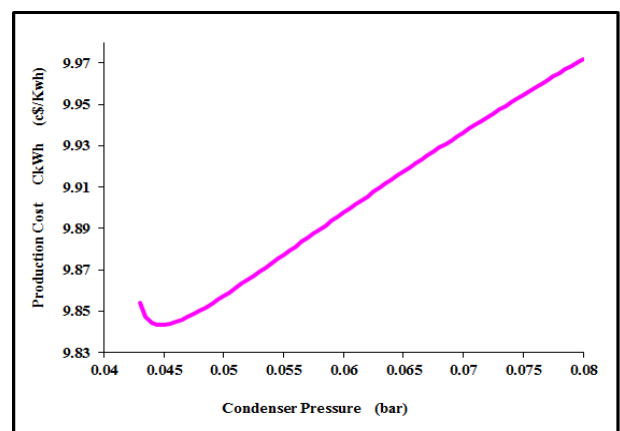


Figure 6. Effect of condenser pressure variations on the production cost of electricity

Figure 6 illustrates the effect of the condenser pressure (P_{Cond}) on the production cost of electricity, which is defined by Eq. 10 for a CCGTs with a triple-pressure HRSG. As seen from that, production cost of electricity decreases with an increase in the P_{Cond} , until it reaches to the optimal value and then increases with further increase in the P_{Cond} . The result shows that the optimal value for the P_{Cond} , at which the minimum production cost of electricity is achieved, is 0.045 bar.

Table 2 shows a comparison of the results between the initial case and optimised case of triple-pressure HRSG. The results show that the financial parameters are significantly better than the initial case. Thermoeconomic optimisation intend to achieve a trade-off between enhance the thermal efficiency of the power plant and economic parameters (B , C_{kWh}).

Table 2. Comparison between the initial case and optimised case of triple-pressure HRSG CCGTs

Parameter	Initial case	Optimised case	
		Objective function: Min. Production cost	Objective function: Max. Annual cash flow
Condenser pressure (P_{Cond})	0.057 bar	0.045 bar	0.0445 bar
Heat transfer area (A_{Cond})	6838 m ²	15201.9 m ²	16454.16 m ²
Combined cycle-efficiency (η_{CCGT})	55.65 %	56.02 %	56.03 %
Combined cycle-gross power (W_{CCGT})	283.07 MW	284.93 MW	285.02 MW
Condenser Cost (C_{Cond})	\$2.091 million	\$4.687 million	\$5.077 million
Production cost (C_{kWh})	9.886 cent/kWh	9.843 cent/kWh	-
Annual cash flow (B)	\$32.153 million	-	\$33.273 million

The results illustrates that the optimisation objectives function (C_{kWh} and B) led to different optimised parameter P_{Cond} . Therefore, it is important to specify the thermoeconomic optimisation objective. In the case of a strong competition in liberalised electricity market and therefore it is difficult to sell the generated energy, the production costs per unit of electrical output (C_{kWh}) optimisation should be done. Otherwise, in a regulated market with fixed energy prices where it is easy to sell the energy, the annual cash flow (B) optimisation is better because this way more incomes are got due to the higher power reached.

In our case study triple-pressure HRSG combined cycle power plant, the efficiency of the selected combined cycle configuration could be increased by about 0.37 % and the electrical output by more than 1.5 MW. On the other hand, the annual cash flow (B) was increased by \$1.1 million by optimal selection of the condenser pressure. In addition, by selecting the optimal condenser pressure the production cost of electricity was decreased by 0.043 cent per kilowatt-hour.

5. CONCLUSION

In this paper, the thermoeconomic optimisation of a triple-pressure HRSG for combined cycle power plants was investigated, and was applied on an example. The analysis indicated that the thermoeconomic optimisation procedure used in this study led to a significant improvement for the economic parameters. Compared with the initial case, the annual cash flow was increased by \$1.1 million, and the production cost of the electricity was decreased by 0.043 cent Kwh.

The results achieved here depend on many considerations such as the gas turbine selection, power plant configurations and the fuel price where the power plant was installed. The authors still work to investigate the effect of using all the operating parameters for the steam condenser to study both the thermodynamic and thermoeconomic optimisation for triple-pressure HRSG combined cycle power plant.

6. REFERENCES

- [1] Alus, M., 2014, "New Method for Thermoeconomic Optimisation of Parameters of Triple-Pressure Combined Cycle Power Plants", Ph.D. thesis, University of Belgrade.
- [2] Lakovic', M. S., Mitrovic', D. D., 2010, "Impact of the Cold End Operating Conditions on Energy Efficiency of the Steam Power Plants", Thermal Science, vol. 14, no. 1, pp. 553-566.
- [3] Wagner, W., Kruse, A., 1998, "Properties of Water and Steam", IAPWS-IF97, Springer, Berlin.
- [4] Haldkar, V., Sharma, A. K., Ranjan, R. K. and Bajpai, V. K., 2013, "Parametric Analysis of Surface Condenser for thermal Power Plant", International Journal of Thermal Technologies, vol. 3, no. 4, pp. 155-159.
- [5] Ozdamar, G., Ozturk, R., 2012, "Optimisation of A condenser in A thermal Power Plant", 9th International Conference on Heat Transfer, Fluid Mechanics and Thermodynamics.
- [6] Behr, H. D., Diederichsen, C., 1988, "Calculation equations for enthalpy and entropy of the components of air and combustion gases", BWK, vol. 40, no. 1-2, pp. 30-33.
- [7] Li, W., Priddy, P., 1985, "Power Plant System Design", John Wiley & Sons, Canada.
- [8] Roosen, P., Uhlenbruck, S. and Lucas, K., 2003, "Pareto optimisation of a combined cycle power system as a decision support tool for trading off investment vs. operating costs", International Journal of Thermal Sciences, vol. 42, p. 553-560.
- [9] Behbahani-nia, A., Sayadi, S. and Soleymani, M., 2010, "Thermoeconomic optimisation of the pinch point and gas-side velocity in heat recovery steam generators", Journal of Power and Energy, vol. 224, pp. 761-771.
- [10] Silveira, J., Tuna, C., 2003, "Thermoeconomic analysis method for optimisation of combined heat and power systems, Part I", Progress in Energy and combustion Science, vol.29, pp. 479-485.
- [11] Attala, L., Facchini, B. and Ferrara, G., 2001, "Thermoeconomic optimisation method as design tool in gas-steam combined plant realization", Energy Conversion and Management, vol. 42, pp. 2163-2172.
- [12] Valdes, M., Duran, and Rovira, A., 2003, "Thermoeconomic optimisation of combined cycle gas turbine power plants using genetic algorithms", Applied Thermal Engineering, vol. 23, pp.2169-2182.

- [13] Manuel, V., Rovira A. and Durán, M. D., 2004, "Influence of the heat recovery steam generator design parameters on the thermoeconomic performance of combined cycle gas turbine power plants", International Journal Energy Research, vol. 28, pp. 1243-1254.

7. Appendix

Nomenclature			
A	Heat transfer area (m^2)	η	efficiency
c	Specific cost (\$/kWh)	C	Compressor
C	Cost (\$)	CC	Combustion chamber
h	Specific enthalpy (kJ/kg)	CCGT	Combined cycle power plant
h	Number of operating hours of the plant per year	Cond.	Condenser
H_l	Lower heat value of fuel (kJ/kg)	f	Fuel
I_{tot}	Total income (\$/year)	HRSG	Heat recovery steam generator
\dot{m}	Mass flow rate (kg/s)	HP	High pressure

N	Economic life of the plant (Year)	IP	Intermediate pressure
k	Specific cost of components for HRSG (\$)	LP	Low pressure
P	Pressure (bar)	E	Economiser
\dot{Q}	Heat recovered (kW)	V	Evaporator
R	Additional cost factor	SH	Superheater
S	Selling price of 1 kWh (\$/kWh)	RE	Reheat
T	Temperature (K or $^{\circ}C$)	ST	Steam turbine
W	Power (MW)		

Image Forgery Detection Based On Sets of 9 Dimensional Convex Equations

Vinoth S
 ECE Department
 IIT Hyderabad
 India

Abstract: Image forgery creation is analyzed from 9 dimensional convex equation analysis

Keywords: 9 dimensions, convex equation

1. INTRODUCTION

Image forgery is created based on 9 dimensional convex equations and their analysis.

2. EQUATIONS AND THEIR ANALYSIS

A 18 dimensional plane is considered. Bases are taken for each of the dimensions as [1 0 0 0 ... 0], [0 1 0 0 0 ... 0], [0 0 1 0 0 0 ... 0] ... [0 0 0 ... 1]. Equations are considered by clubbing two bases in terms of a parabola with the ends symmetric to origin. Likewise 9 such combinations are considered to achieve 9 such convex shapes centered at origin.

Equation of 1st convex shape

$Y_1 = a_1[1\ 0\ 0\ 0\ \dots\ 0] + b_1[0\ 1\ 0\ 0\ 0\ \dots\ 0]$ where a_1 and b_1 are vectors.

$Y_2 = a_2[0\ 0\ 1\ 0\ 0\ 0\ \dots\ 0] + b_2[0\ 0\ 0\ 1\ 0\ 0\ 0\ \dots\ 0]$ where a_2 and b_2 are vectors.

$Y_3 = a_3[0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ \dots\ 0] + b_3[0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ \dots\ 0]$ where a_3 and b_3 are vectors.

.

.

.

$Y_9 = a_9[0\ 0\ 0\ \dots\ 1\ 0] + b_9[0\ 0\ 0\ \dots\ 0\ 1]$ where a_9 and b_9 are vectors.

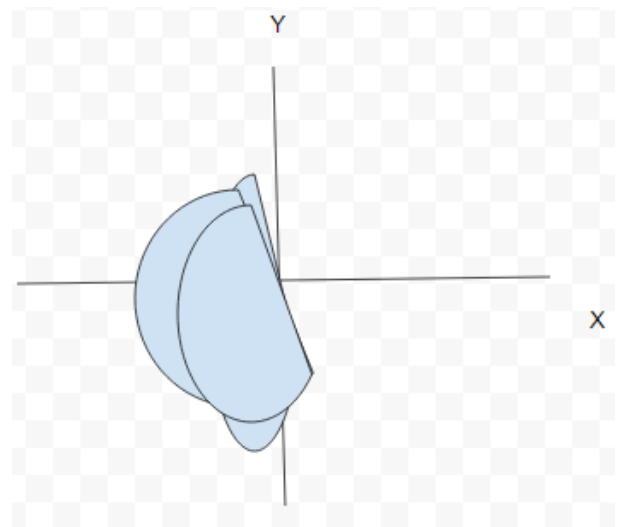


Fig. 1 Y_1 , Y_2 and Y_3

3. CREATION OF FORGED IMAGE

3.1 Projection of considered image

The image equation is calculated based on backpropagation neural network's weight vectors as seen in reference [2].

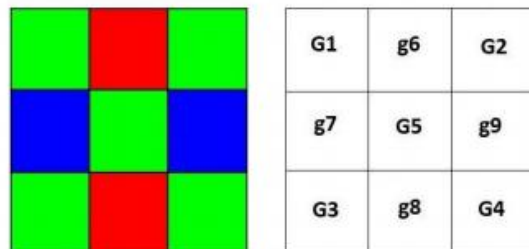


Fig. 2 Calculation of image equation(a)

From Fig 2 [G1 G2 G3 G4 G5] are given as input and [g6 g7 g8 g9] are given as output in Fig 3.

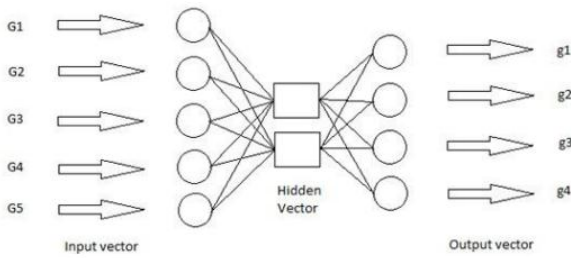


Fig. 3 Calculation of image equation(b)

The neural networks are trained as seen above and the average of all weights stacked together forms the image equation. Average of all neural network weights is taken as the equation of the image considered. This equation is projected on the 9 different convex shapes considered above i.e. $Y_1 Y_2 Y_3 \dots Y_9$.

3.2 Detection of forged image

Equations whose null spaces of projections on the 9 convex shapes are taken. If there is an intersection of this equation with that of the image equation the image is not forged.

Equations whose null spaces of projections on the 9 convex

shapes are taken. If there is an intersection of this equation with that of the image equation the image is not forged. This analysis is repeated for all the null spaces arrived for at the convex equational base. This is extended to one more dimension to seek whether the image is forged.

Projections are taken from this equation onto the convex shapes on these convex equations to create a forged image.

4. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

5. REFERENCES

- [1] Vinoth S, Image forgery analysis based on 9 dimensional convex equations
- [2] Vinoth S, ES Gopi, Neural Network modelling of color array filter for image forgery detection using kernel LDA
- [3] Vinoth S, Image forgery detection based on shape of eye ball
- [4] Vinoth S, Image forgery detection based on 9 dimensional convex equations

New Physiologically Active Kaempferol Glucoside from *Abutilon pannosum*.

Abdullah I. Kamel
Chemistry Department
Faculty of Science
Mansoura University
Mansoura, Egypt

Manal G. Elfedawy
Chemistry Department
Faculty of Science
Mansoura University
Mansoura, Egypt

Maha M.ElShamy
Botany Department
Faculty of Science
Mansoura University
Mansoura, Egypt |

Mamdouh Abdel-Mogib
Chem. Dept., Fac. of Sci.,
Mansoura University
Mansoura, Egypt

Abstract:

The phytochemical investigation of *Abutilon pannosum* (Malvaceae) collected from arable land of Tennidah village, Dakhla Oasis, Western Desert, Egypt, resulted in the separation and identification of a new flavonoid, kaempferol 4'-O-(6''-O-E-p-coumaroyl)- β -D-glucopyranoside **1** from butanol fraction, in addition to the identification of the volatile constituents of petroleum ether and methylene chloride by GC/MS analysis. Structure of the separated compound was elucidated by spectral analysis. Additionally, the antimicrobial, antioxidant activities and cytotoxicity of different fractions of *Abutilon pannosum* and compound **1** were evaluated. The antimicrobial activity indexes of methanol extract (Ap C), compound **1**, methylene chloride extract (Ap B) and petroleum ether extract (Ap A) against *Escherichia coli* were found to be 66.7%, 50.0%, 33.3%, and 0.0%, respectively. While the activity indexes of compound **1**, (Ap B), (Ap C), (Ap A) against *staphylococcus aureus* were 68.2%, 63.6%, 59.1% and 9.1%, respectively, and against *Candida albicans*, their activity indexes were 38.8%, 38.5%, 34.6%, and 11.5%, respectively. The radical scavenging activity of the extracts and standard were found to be in the following order: ascorbic acid (88.6%) > (Ap C) (84.4%) > compound **1** (76.8%) > (Ap B) (59.2%) > (Ap A) (6.3%). The IC50 values against HePG2 indicated that the cytotoxicity of extracts decreased in the order: compound **1** was "very strong", (Ap C) was "strong", (Ap B) was "moderate", and (Ap A) was "weak", while against PC3 of compound **1** was "strong", (Ap B and Ap C) were "moderate", and (Ap A) was weak, and against MCF-7 compound **1** and (Ap C) were "strong", (Ap B) was "moderate", and (Ap A) was "non-cytotoxic".

Keywords: Malvaceae, *Abutilon pannosum*, flavonoid, kaempferol 4'-O-(6''-O-E-p-coumaroyl)- β -D-glucopyranoside, antimicrobial activity, radical scavenging activity, cytotoxicity.

1.1. INTRODUCTION

The Malvaceae, or the mallows, are a family of flowering plants estimated to contain 243 genera with 4225+ species [1]. Well-known members of this family include okra, cotton, and cacao. The genus *Abutilon* is one of the larger genera of the family Malvaceae [2]. Many authors recorded *A. pannosum* in Eastern Desert of Egypt [3, 4]. *A. pannosum* (G.Forst.) Schlttdl. was recorded at edges of cultivation and arable lands in some oasis of Southern part of Western Desert [5]. On March 2014, it was collected from, Esbet-Qattara at El-Monira village nearby Pharous Temple, Kharga Oasis (N 25°38'38.64" -E 30°38'38.88") and at Tennidah village (N 25°30'31.08" - E 29°19'57.84") in Dakhla Oasis as a part of Nubian Sothern part of Western Desert, Egypt, along road sides and arable lands. Plantations and other cultivated areas are the preferred habitats of *A. pannosum* [5]. *A. pannosum* is a Chaemophyte, Perennial shrubby herb, 1 - 3 m, with erect or spreading stems branch out from the base. The stems are covered with yellow

hairs. The heart-shaped leaves are velvety on both surfaces and have irregularly toothed edges. The flowers are located on the ends of long stalks, and are either solitary or found in clusters. Each flower has five petals which are approximately two centimeters long, and are either yellow or orange with a dark red to purple base. It produces small, rounded fruits which have obvious ridges and furrows. Each fruit is densely packed with flat mericarps. The brown, hairy seeds are usually kidney-shaped. The flowering period is between January and June [6]. *A. pannosum* is found in tropical Africa, including Djibouti, Eritrea, northern Ethiopia, Egypt and Sudan, and across to West Africa and the Cape Verde Islands. This species also occurs in Asia, where it is found in Saudia Arabia, the United Arab Emirates, Pakistan and India. In Pakistan, it is restricted to Sindh [7].

Preliminary phytochemical screening of *A. pannosum* showed the presence of Alkaloids, cardiac glycosides and steroids [8],

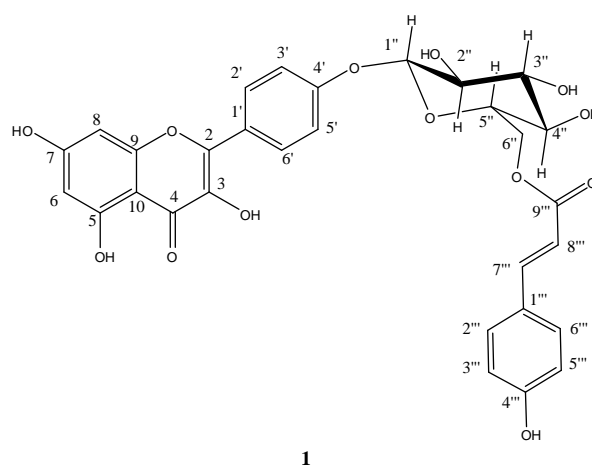
seven known compounds namely, lupeol, beta-sitosterol, stigmasterol, methyl-4-hydroxybenzoate, taraxacin, ursolic acid, and beta-sitosterol-3-O-beta-D-glucopyranoside, have been isolated from the EtOAc soluble fraction of *A. pannosum* [9].

In this article, we present the results of the phytochemical reinvestigation, as well as the antimicrobial, antioxidant and antitumor activities of *Abutilon pannosum*.

2. RESULTS AND DISCUSSION

2.1 Phytochemical evaluation

Separation of butanol extract (Ap D) of *Abutilon pannosum* afforded a new natural product, kaempferol 4'-O-(6''-O-E-*p*-coumaroyl)-β-D-glucopyranoside **1**. Additionally, the petroleum ether (Ap A) and methylene chloride (Ap B) fractions were analyzed by GC/MS. A sample from (Ap A) extract afforded 31 compound, with n-nonadecane (8.28%), n-eicosane (7.63%), n-octadecane (7.31%), n-heptadecane (7.29%), and n-hexadecane (6.54%) being the major components. A sample from (Ap B) extract afforded 29 compound, with phytol (31.69%), squalene (15.76%), neophytadiene (5.58%), and n-hexadecanoic acid (4.97%) being the major components. NMR guided chromatographic separation led to two promising sub-fractions that were analyzed by GC/MS. A petroleum ether sub-fraction (Ap a) afforded 15 compound, with methyl palmitate (38.98%) as a major component, and a methylene chloride sub-fraction (Ap b) afforded 9 compounds with squalene (61.71%) as a major component. Compound **1** was identified by spectral analysis. The ¹H NMR spectrum (Table 1) showed a signal pattern of a kaempferol derivative with H-6, H-8 signals of ring A at δ 6.21, 6.39, in agreement with 5,7-dihydroxy [10, 11], and AA'BB' spin system of ring B with coupling of 8.4 Hz at δ 7.97 and 6.79, indicating the probable 4'-substitution. The spectrum showed also signals of a glucose unit, with the anomeric proton signal at δ 5.46 as a doublet with J of 7.2 Hz, indicating a β-D-glucopyranoside. The signals of H-6 protons shifted downfield at δ 4.05, 4.28 indicating the probable esterification to *p*-coumaroyl that gave AA'BB' spin system at δ 6.86, 7.38 and the ethylenic H-7, H-8 at δ 6.12, 7.35 as a pair of doublets with trans coupling of 16 Hz. The probable glycosidation at C-3-OH was excluded on the bases of notable shifts of ring B signals [10, 11]. Thus, **1** was identified as kaempferol 4'-O-(6''-O-E-*p*-coumaroyl)-β-D-glucopyranoside, rather than its isomer cephaeoside, isolated previously from the same species [9]. The proposed structure for **1** was confirmed by 1- and 2-D NMR analysis (DEPT, HSQC, HMBC)



2.1. Biological applications

2.1.1. Antimicrobial activity assessment

The antimicrobial activity indexes of methanol extract (Ap C), compound **1**, methylene chloride extract (Ap B) and petroleum ether extract (Ap A) were presented in Table 2 against *Escherichia coli* were found to be 66.7%, 50%, 33.3%, and 0.0%, respectively. While the activity indexes of compound **1**, (Ap B), (Ap C), and (Ap A) against *staphylococcus aureus* were 68.2%, 63.6%, 59.1% and 9.1%, respectively, and against *Candida albicans*, their activity indexes were 38.8%, 38.5%, 34.6%, and 11.5%, respectively (table 2).

2.1.2 Free radical scavenging activity assessment

The anti-oxidant activities of the extracts of *A. pannosum* were presented in Table 3 by (ABTS) method [12, 13, 14]. Methanol extract (Ap C) had the highest scavenging activity. The radical scavenging activity of the extracts and standard decreased in the following order: ascorbic acid, methanol (Ap C), compound **1**, methylene chloride (Ap B) and petroleum ether (Ap A) (table 3), respectively.

Table3: ¹H, ¹³C NMR and HMBC data of compound **1**

Position	¹ H (multiplicity, J)	¹³ C	DEPT	Long range coupled protons, from HMBC
2	-----	160.44	C	H-2',H-6'
3	-----	133.49	C	-----
4	-----	177.85	C	-----
5	-----	156.81	C	-----
6	6.21, (1H, br.s)	99.23	CH	-----
7	-----	164.64	C	-----
8	6.39, (1H, br.s)	94.13	CH	-----
9	-----	156.90	C	H-8
10	-----	104.31	C	H-6
1'	-----	125.36	C	H-3',H-5'
2' & 6'	7.97, (2H, d, J= 8.4 HZ)	131.28	CH	-----
3' & 5'	6.79, (2H, d, J= 8.4 HZ)	116.21	CH	-----
4'	-----	160.24	C	H-2',H-6', H-3', H-5'
1''	5.46, (1H, d, J= 7.2 HZ)	101.39	CH	-----
2''	3.37, (1H, m)	74.67	CH	-----
3''	3.28, (1H, m)	76.65	CH	-----
4''	3.24, (1H, m)	70.41	CH	-----
5''	3.21, (1H, m)	74.57	CH	-----
6''a	4.28, (1H, br.d, J= 12 HZ)	63.40	CH ₂	-----
6''b	4.05, (1H, dd, J= 12, 5.2 HZ)			
1'''	-----	121.21	C	H-2''',H-6'''
3''' & 5'''	6.86, (2H, d, J= 8.4 HZ)	115.54	CH	H-3''',H-5'''
2''' & 6'''	7.38, (2H, d, J= 8.4 HZ)	130.61	CH	H-3''',H-5''', H-7'''
4'''	-----	161.59	C	H-2''',H-6''',H-3''',H-5'''
7'''	7.35, (1H, d, J= 16 HZ)	145.06	CH	H-3''',H-5'''
8'''	6.12, (1H, d, J= 16 HZ)	114.07	CH	-----
9'''	-----	166.63	C	H-7''',H-8'''

Table 2: The inhibition zone in mm of the extracts of *Abutilon pannosum* and compound **1** compared to standard antibiotics [15]

Compound / Fraction	<i>E. coli</i> (mg/ml)		<i>S. aureus</i> (mg/ml)		<i>C. Albicans</i> (mg/ml)	
	Diameter of inhibition zone (in mm)	% Activity index	Diameter of inhibition zone (in mm)	% Activity index	Diameter of inhibition zone (in mm)	% Activity index
1	12	50	15	68.2	10	38.8
Ap B	8	33.3	14	63.6	10	38.5
Ap C	16	66.7	13	59.1	9	34.6
Ap A	NA	----	2	9.1	3	11.5
Ampicillin	24	100	22	100	NA	----
Colitrimazole	NA	----	NA	----	26	100

NA → No Activity.

Table 3: Antioxidant activity of the extracts of *Abutilon pannosum* and compound **1** by ABTS

Compound / Fraction	Absorbance of samples	% Inhibition
Control of ABTS	0.520	0%
Ascorbic-acid	0.059	88.6%
Ap C	0.081	84.4%
1	0.120	76.8
Ap B	0.212	59.2%
Ap A	0.487	6.3%

The IC50 values against HePG2 indicated that the cytotoxicity of extracts decreased in the order: compound **1** was "very strong", (Ap C) was "strong", (Ap B) was "moderate", and (Ap A) was "weak", while against PC3 of compound **1** was "strong", (Ap B and Ap C) were "moderate", and (Ap A) was "weak", and against MCF-7 of compound **1** and (Ap C) were "strong", (Ap B) was "moderate", and (Ap A) was "non-cytotoxic" (table4).

Table 4: Cytotoxic activity assessment of *Abutilon pannosum* extracts and compound **1** against human tumor cells HePG2, MCF-7 and PC3 [16, 17, 18]

• IC50(μg/ml): 1 – 10 (very strong). 11 – 20 (strong). 21 – 50 (moderate). 51 – 100 (weak) and above 100 (non-cytotoxic).

2.2.3 Cytotoxic activity assessment

Compound / Fraction	In vitro Cytotoxicity IC50 (µg/ml)•		
	HePG2	MCF-7	PC3
5-FU	6.6±0.24	4.7±0.11	9.6±0.27
1	8.6±0.73	14.0±0.89	19.8±1.13
Ap C	10.4±0.45	15.3±0.77	23.6±1.06
Ap B	49.5±2.34	33.6±2.12	22.3±0.98
Ap A	80.4±4.58	>100	90.7±5.47

• 5-FU = 5-fluorouracil

3. EXPERIMENTAL

3.1 Instrumentations

3.1.1 ¹H-NMR

The NMR spectra were recorded in deuterated chloroform (CDCl₃) or Dimethyl Sulfoxide-d₆ (DMSO-d₆) at either Faculty of Pharmacy, Benisweef University, on Bruker Avance III 400 MHz for ¹H and 100 MHz for ¹³C (Bruker AG, Switzerland) with BBFO Smart Probe and Bruker 400 AEON Nitrogen-Free Magnet, and Topspin 3.1 software for data analysis, or at Faculty of Science, Cairo University, on Varian Mercury VX-300 NMR spectrometer, run at 300 MHz for ¹H and at 75.46 MHz for ¹³C. Chemical shifts are quoted in δ (ppm) and were related to that of the solvents.

3.1.2 GC/MS analysis

Method 1: GC/MS analysis was performed at the Central Laboratory of the Ministry of Agriculture, Al Bhooth Str, Cairo, on Agilent 6890 gas chromatograph equipped with an Agilent mass spectrometric column HP-5ms (30 m x 0.32 mm x 0.25 µm film thickness). Samples were injected under the following condition: Helium was used as carrier gas at approximately 1 ml /min, pulsed splitless mode. The solvent delay was 3 min and the injection size was 1.0 µl. The mass spectrophotometric detector was operated in electron impact ionization mode with an ionizing energy of 70 eV, scanning from m/z 50 to 500. The ion source temperature was 230°C and the quadruple temperature was 150°C. The electron multiplier voltage (EM voltage) was maintained at 1250 v above auto tune. The instrument was manually tuned using perfluorotributyl amine (PFTBA). The GC temperature program was started at 60°C then elevated to 280°C at rate of 8°C/min. and 10 min. hold at 280°C the detector and injector temperature were set at 280°C and 250°C, respectively. Wiley and Nist 05 mass spectral database was used in the identification of the separated peaks.

Method 2: GC/MS analysis was performed at the National Research Center (NRC), Dokki, Cairo, and on a varian GC interfaced to Finnigan SSQ 7000 Mass Selective Detector (MSD) with ICIS V2.0 data system for MS identification of the GC components. The column used was DB-5 (J & W Scientific, Folsom, CA) cross-linked fused silica capillary column (30 m long, 0.25 mm internal diameter) coated with poly dimethyl siloxane (0.5 µm film thickness). The oven temperature was programmed from 50°C for 3 min., at isothermal, then heating by 7°C / min. to 250°C and isothermally for 10 min., at 250 °C. Injector temperature was 200°C and the volume injected was 0.5 µl. Transition-line and ion source temperature were 250°C and 150°C respectively. The mass spectrometer had a delay of

3 min. to permit the solvent plead and then scanned from m/z 50 to m/z 300. Ionization energy was set at 70 ev.

3.2 Materials

3.2.1 Plant material

Abutilon pannosum was collected from Tennidah village, Dakhla oasis, and new valley governorate in western desert on March 2014 by Dr. Maha Elshamy, Botany Department, Faculty of Science, Mansoura University. The leaves were used for investigation.

3.2.2 Cell line

Hepatocellular carcinoma (liver) HePG-2, mammary gland (breast) MCF-7 and Human (prostate) cancer cell line PC3 were obtained from ATCC via Holding company for biological products and vaccines (VACSERA), Cairo, Egypt.

3.2.3 Chemical reagents and solvents

The reagents RPMI-1640 medium, MTT, DMSO, 5-fluorouracil (Sigma co., St. Louis, USA) and Fetal Bovine serum (GIBCO, UK) were used for cytotoxicity assessment. Thin layer chromatography and preparative (TLC) were performed on silica gel (Kieselgel 60, GF 254). Silica gel (60-120 Mesh) for column chromatography (Nice Chemicals Co.), Lipophilic sephadex LH-20 for column was obtained from sigma chemical company, hexane was obtained from Alpha Chemika, methylene chloride was obtained from SDFCL SD fine-chem limited; EtOAc, MeOH, BuOH, C₆H₆, (CH₃)₂CO and Na₂SO₄, were obtained from Adowic.

3.2.4 Spray reagent

P-Anisaldehyde-sulphuric acid reagent was prepared freshly by adding conc. sulphuric acid (8 µl) to solution of p-Anisaldehyde (0.5 µl) in glacial acetic acid (10 µl) and methanol (85 µl). The chromatogram after spraying was heated at 100°C until the spots attained maximum color (Wagner et al., 1984).

3.3 Processing of *Abutilon pannosum* leaves

The freshly collected plant material of *Abutilon pannosum* was cut into small pieces and air dried in shade at room temperature. After drying the plant material, leaves were ground to give (400.21 g) of dried powdered material, and extracted by soxhlet extractor using different solvents of different polarity; petroleum ether 60-80, methylene chloride and methanol respectively to attain three fractions: petroleum ether fraction (Ap A, 10.920 g), Methylene chloride fraction (Ap B, 3.355 g), and Methanol fraction (Ap C, 14.974 g).The Butanol fraction (Ap D, 7.043g) separated from Methanol fraction.

3.3.1 Processing of petroleum ether fraction (Ap A)

A sample from Ap A fraction (0.816 g) was analyzed by GC/MS using method 1. Then, the fraction Ap A (10.104 g) was defatted through dissolving in cold methanol to obtain defatted part (1.752 g). A part of which (1.152 g) was subjected to column chromatography over silica gel using a series of eluents from petroleum ether / ethyl acetate combinations of

increase polarity. Ninety-five sub fraction were collected from the column and by inspecting their TLC, five sub-fractions (14a, 17a, 78a, 85a, and 90a) were chosen to ¹H-NMR, and according to the results of ¹H-NMR, the sub fraction (14a, 0.317 g) was chosen for GC/MS analysis by method 2 as (Ap a). The remaining part (0.513 g) was used for biological activity.

3.3.1.1 Chemical composition of fraction (Ap A) by GC/MS

Petroleum ether Fraction Ap A afforded by GC/MS using method 1 n-undecane (2) (R_t= 9.22 min., 0.55%), n-dodecane (3) (R_t= 11.18 min., 0.57%), 2-methyldodecane (21) (R_t= 12.34 min., 0.25%), 2,6-dimethyldodecane (24) (R_t= 12.49 min., 0.44%), n-tridecane (4) (R_t= 12.99 min., 0.165%), n-tetradecane (5) (R_t= 14.69 min., 2.62%), phytane (43) (R_t= 15.68 min., 1.65%), n-pentadecane (6) (R_t= 16.28 min., 4.79%), n-hexadecane (7) (R_t= 17.78 min., 6.54%), 2-methyl hexadecane (22) (R_t= 18.67 min., 1.97%), n-heptadecane (8) (R_t= 19.20 min., 7.29%), 1-methyl-9H-fluorine (53) (R_t= 19.49 min., 1.79%), n-octadecane (9) (R_t= 20.55 min., 7.31%), 3-methyloctadecane (23) (R_t= 21.35 min., 3.75%), n-nonadecane (10) (R_t= 21.83 min., 8.28%), n-hexadecanoic acid (25) (R_t= 22.67 min., 5.78%), n-eicosane (11) (R_t= 23.05 min., 7.63%), n-heineicosane (12) (R_t= 24.21 min., 5.39%), phytol (46) (R_t= 24.41 min., 3.09%), methyl linolenate (31) (R_t= 24.76 min., 4.15%), n-docosane (13) (R_t= 25.33 min., 4.69%), 14-β-H-pregnane (45) (R_t= 25.54 min., 1.68%), n-tricosane (14) (R_t= 26.40 min., 3.63%), n-tetracosane (15) (R_t= 27.43 min., 3.21%), n-pentacosane (16) (R_t= 28.41 min., 3.34%), n-hexacosane (17) (R_t= 29.36 min., 1.95%), n-heptacosane (18) (R_t= 30.28 min., 1.40%), n-octacosane (19) (R_t= 31.21 min., 0.79%), squalene (50) (R_t= 31.60 min., 2.78%), n-nonacosane (20) (R_t= 32.27 min., 0.70%) and vitamin E (51) (R_t= 35.85 min., 0.50%).

3.3.1.2 Chemical composition of sub-fraction (Ap a) by GC/MS

Petroleum ether sub-fraction (Ap a) afforded by GC/MS using method 2 methyl octadeca-12,15-dienoate (30) (R_t= 30.93 min., 2.82%), methyl octadeca-10,13-dienoate (29) (R_t= 32.87 min., 1.61%), n-octadecane (9) (R_t= 36.64 min., 6.26%), n-nonadecane (10) (R_t= 39.11 min., 8.48%), methyl palmitate (28) (R_t= 39.83 min., 38.98%), 2-[4-methyl-6-(2,6,6-trimethylcyclohex-1-enyl)hexa-1,3,5-trienyl]cyclohex-1-en-1-carboxaldehyde (48) (R_t= 40.00 min., 1.21%), n-eicosane (11) (R_t= 41.47 min., 8.28%), n-heineicosane (12) (R_t= 43.76 min., 8.08%), methyl linolenate (31) (R_t= 43.87 min., 5.65%), methyl stearate (32) (R_t= 44.42 min., 1.61%), n-docosane (13) (R_t= 45.93 min., 5.85%), n-tricosane (14) (R_t= 48.02 min., 4.24%), n-tetracosane (15) (R_t= 50.03 min., 3.23%), n-pentacosane (16) (R_t= 51.96 min., 2.22%) and n-hexacosane (17) (R_t= 53.83 min., 1.41%).

3.3.2 Processing of methylene chloride fraction (Ap B)

A sample from Ap B fraction (0.683 g) was analyzed by GC/MS using method 1, a part from it (2.005 g) was subjected

to column chromatography over silica gel using a series of eluents from hexane / ethyl acetate combinations of increase polarity. Fifty-eight sub-fraction were collected from the column and by inspecting their TLC, nine sub fractions (5b, 7b, 12b, 16b, 20b, 25b, 35b, 41b and 43b) were chosen to ¹H-NMR, and according to the results of ¹H-NMR, the sub-fractions (5b, 7b, 0.483 g) were mixed together as (Ap b) and were chosen for GC/MS analysis by method 2. The samples (16b, 25b) were mixed together and subjected to reversed phase column according to ¹H-NMR results but no any benefit results were obtained. The remaining part (0.497 g) was used for biological activity.

3.3.2.1 Chemical composition of fraction (Ap B) by GC/MS

Methylene chloride fraction Ap B afforded by GC/MS using method 1 n-tetradecane (5) (R_t= 14.67 min., 0.33%), n-pentadecane (6) (R_t= 16.26 min., 0.85%), dehydroxyloliolide (33) (R_t= 17.08 min., 0.39%), n-hexadecane (7) (R_t= 17.75 min., 1.6%), crocetane (43) (R_t= 18.45 min., 0.82%), 2-methyl hexadecane (22) (R_t= 18.65 min., 0.28%), n-heptadecane (8) (R_t= 19.16 min., 1.85%), pristane (40) (R_t= 19.25 min., 1.36%), loliolide (35) (R_t= 20.42 min., 1.01%), n-octadecane (9) (R_t= 20.51 min., 2.64%), neophytadiene (42) (R_t= 21.04 min., 5.58%), 2,6,10-trimethyldodecane (38) (R_t= 21.13 min., 1.06%), n-nonadecane (10) (R_t= 21.79 min., 2.32%), methyl palmitate (28) (R_t= 22.12 min., 1.36%), n-hexadecanoic acid (25) (R_t= 22.61 min., 4.97%), n-eicosane (11) (R_t= 23.01 min., 3.27%), n-heneicosane (12) (R_t= 24.18 min., 2.73%), methyl linolenate (31) (R_t= 24.26 min., 0.43%), phytol (46) (R_t= 24.39 min., 31.69%), 9-octadecenoic acid (26) (R_t= 24.73 min., 2.75%), n-docosane (13) (R_t= 25.30 min., 2.79%), n-tricosane (14) (R_t= 26.37 min., 2.54%), 4,8,12,16-tetramethyl heptadecan-1,4-olide (47) (R_t= 27.03 min., 1.29%), geranyl acetone (34) (R_t= 27.33 min., 0.55%), n-tetracosane (15) (R_t= 27.40 min., 3.04%), n-pentacosane (16) (R_t= 28.39 min., 2.24%), squalene (50) (R_t= 31.58 min., 15.76%), vitamin E (51) (R_t= 35.58 min., 3.00%) and gibberellin A3 (49) (R_t= 38.41 min., 1.31%).

3.3.2.2 Chemical composition of sub-fraction (Ap b) by GC/MS

Methylene chloride-sub fraction (Ap b) afforded by GC/MS using method 2 4(15),6-cadinadiene (37) (R_t= 24.46 min., 1.76%), trans-β-caryophyllene (36) (R_t= 25.83 min., 5.67%), n-hexadecane (7) (R_t= 27.13 min., 4.44%), phytol (46) (R_t= 37.57 min., 1.82%), 6,10,14-trimethyl-2-pentadecanone (41) (R_t= 37.75 min., 13.27%), methyl 14-methylpentadecanoate (27) (R_t= 39.91 min., 4.63%), dioctyladipate (54) (R_t= 50.11 min., 4.88%), squalene (50) (R_t= 57.94 min., 61.71%) and 3,4,3',4'-tetrahydrospirilloxanthin (52) (R_t= 60.99 min., 1.76%).

3.3.3 Processing of methanol fraction (Ap C)

Methanol fraction (Ap C, 13.651 g) was portioned between water and butanol in separating funnel. The butanol layer was evaporated to dryness affording (ApD, 7.043 g). A sample from (ApD, 3.907 g) was subjected to silica gel column using methanol / methylene chloride eluents of increase polarity. The

sub-fractions were subjected to sephadex LH-20 column using methanol solvent as eluent, followed by TLC purification with eluent system ethyl acetate (9.1 ml) / methanol (0.6 ml) / water (0.3 ml) to give **1** (594 mg, R_f = 0.65).

3.4 Biological applications

3.4.1 Antimicrobial activity assessment

The three fractions and the isolated compound were individually tested against a panel of gram positive (*Staphylococcus aureus*), gram negative (*Escherichia coli*) bacterial and the fungus (*Candida albicans*). Each of the fractions and the isolated compound was dissolved in DMSO and solution of the concentration 1 mg / ml were prepared separately. Paper discs of Whatman filter paper were prepared with standard size (5 mm diameter), were cut and sterilized in

an autoclave. The paper discs were soaked in the desired concentration of the extract solution and placed aseptically in the Petri-dishes containing nutrient agar media (agar 20 g + beef extract 3 g + peptone 5 g) seeded with *Staphylococcus aureus*, *Escherichia coli* and *Candida albicans*. The Petri dishes were incubated at 36°C and the inhibition zones were recorded after one day of incubation. Each treatment was replicate 3 times. The antibacterial activities of a common standard antibiotics *ampicillin*, *gentamicin* and *amphotericin B* were also recorded using the same procedure as above at the same concentration and solvents. The % activity index for each treatment was calculated by the formula:

$$\% \text{ Activity Index} = \frac{\text{Zone of inhibition by test extract (diameter)}}{\text{Zone of inhibition by standard (diameter)}} \times 100$$

Table 5: The MS data of compounds identified by GC/MS

Compound name and number	MS Data :m/z [identity](rel. abund.%)
2 n-undecane	156 [M ⁺] (5.0%), 127 [M-C ₂ H ₅] ⁺ (1.7%), 113 [M-C ₃ H ₇] ⁺ (3.3%), 99 [M-C ₄ H ₉] ⁺ (6.6%), 85 [M-C ₅ H ₁₁] ⁺ (33.3%), 71 [M-C ₆ H ₁₃] ⁺ (53.3%), 57 [M-C ₇ H ₁₅] ⁺ (100%)
3 n-dodecane	170 [M ⁺] (5.0%), 141 [M-C ₂ H ₅] ⁺ (1.0%), 127 [M-C ₃ H ₇] ⁺ (1.7%), 113 [M-C ₄ H ₉] ⁺ (4.0%), 99 [M-C ₅ H ₁₁] ⁺ (6.7%), 85 [M-C ₆ H ₁₃] ⁺ (37.3%), 71 [M-C ₇ H ₁₅] ⁺ (60.0%), 57 [M-C ₈ H ₁₇] ⁺ (100%)
4 n-tridecane	184 [M ⁺] (4.3%), 141 [M-C ₃ H ₇] ⁺ (3.3%), 127 [M-C ₄ H ₉] ⁺ (3.0%), 113 [M-C ₅ H ₁₁] ⁺ (4.0%), 99 [M-C ₆ H ₁₃] ⁺ (8.3%), 85 [M-C ₇ H ₁₅] ⁺ (44.3%), 71 [M-C ₈ H ₁₇] ⁺ (67.7%), 57 [M-C ₉ H ₁₉] ⁺ (100%)
5 n-tetradecane	198 [M ⁺] (3.3%), 173 (0.3%), 160 (1.3%), 145 (3.3%), 127 [M-C ₃ H ₇] ⁺ (3.3%), 113 [M-C ₆ H ₁₃] ⁺ (5.0%), 99 [M-C ₇ H ₁₅] ⁺ (10.3%), 85 [M-C ₈ H ₁₇] ⁺ (50.0%), 71 [M-C ₉ H ₁₉] ⁺ (70.7%), 57 [M-C ₁₀ H ₂₁] ⁺ (100%)
6 n-pentadecane	212 [M ⁺] (4.0%), 183 [M-C ₂ H ₅] ⁺ (0.7%), 169 [M-C ₃ H ₇] ⁺ (1.0%), 155 [M-C ₄ H ₉] ⁺ (2.7%), 141 [M-C ₅ H ₁₁] ⁺ (3.3%), 127 [M-C ₆ H ₁₃] ⁺ (4.0%), 113 [M-C ₇ H ₁₅] ⁺ (6.7%), 99 [M-C ₈ H ₁₇] ⁺ (13.3%), 85 [M-C ₉ H ₁₉] ⁺ (52.3%), 71 [M-C ₁₀ H ₂₁] ⁺ (73.3%), 57 [M-C ₁₁ H ₂₃] ⁺ (100%)
7 n-hexadecane	226 [M ⁺] (3.3%), 197 [M-C ₂ H ₅] ⁺ (0.3%), 182 (3.7%), 170 (13.3%), 155 [M-C ₃ H ₇] ⁺ (16.7%), 141 [M-C ₆ H ₁₃] ⁺ (4.0%), 127 [M-C ₇ H ₁₅] ⁺ (6.7%), 113 [M-C ₈ H ₁₇] ⁺ (9.3%), 99 [M-C ₉ H ₁₉] ⁺ (16.7%), 85 [M-C ₁₀ H ₂₁] ⁺ (52.3%), 71 [M-C ₁₁ H ₂₃] ⁺ (76.7%), 57 [M-C ₁₂ H ₂₅] ⁺ (100%)
8 n-heptadecane	240 [M ⁺] (3.3%), 197 [M-C ₃ H ₇] ⁺ (3.3%), 185[M-C ₄ H ₉] ⁺ (3.3%), 169 [M-C ₅ H ₁₁] ⁺ (6.7%), 141 [M-C ₇ H ₁₅] ⁺ (6.0%), 127 [M-C ₈ H ₁₇] ⁺ (6.7%), 99 [M-C ₁₀ H ₂₁] ⁺ (16.7%), 85 [M-C ₁₁ H ₂₃] ⁺ (60.0%), 71 [M-C ₁₂ H ₂₅] ⁺ (80.0%), 57 [M-C ₁₃ H ₂₇] ⁺ (100%)
9 n-octadecane	254 [M ⁺] (3.3%), 211 [M-C ₃ H ₇] ⁺ (0.7%), 178 (6.7%), 155 [M-C ₇ H ₁₅] ⁺ (3.3%), 141 [M-C ₈ H ₁₇] ⁺ (4.7%), 127 [M-C ₉ H ₁₉] ⁺ (6.7%), 113 [M-C ₁₀ H ₂₁] ⁺ (10.7%), 99 [M-C ₁₁ H ₂₃] ⁺ (18.7%), 85 [M-C ₁₂ H ₂₅] ⁺ (60.0%), 71 [M-C ₁₃ H ₂₇] ⁺ (80.0%), 57 [M-C ₁₄ H ₂₉] ⁺ (100%)
10 n-nonadecane	268 [M ⁺] (3.3%), 197 [M-C ₃ H ₇] ⁺ (2.6%), 141 [M-C ₉ H ₁₉] ⁺ (6.3%), 127 [M-C ₁₀ H ₂₁] ⁺ (8.0%), 113 [M-C ₁₁ H ₂₃] ⁺ (12.3%), 99 [M-C ₁₂ H ₂₅] ⁺ (16.7%), 85 [M-C ₁₃ H ₂₇] ⁺ (62.7%), 71 [M-C ₁₄ H ₂₉] ⁺ (80.0%), 57 [M-C ₁₅ H ₃₁] ⁺ (100%)
11 n-eicosane	282 [M ⁺] (3.3%), 197 [M-C ₆ H ₁₃] ⁺ (3.3%), 155 [M-C ₉ H ₁₉] ⁺ (4.0%), 141 [M-C ₁₀ H ₂₁] ⁺ (6.7%), 127 [M-C ₁₁ H ₂₃] ⁺ (9.3%), 113 [M-C ₁₂ H ₂₅] ⁺ (13.0%), 99 [M-C ₁₃ H ₂₇] ⁺ (20.0%), 85 [M-C ₁₄ H ₂₉] ⁺ (64.0%), 71 [M-C ₁₅ H ₃₁] ⁺ (80.3%), 57 [M-C ₁₆ H ₃₃] ⁺ (100%)
12 n-heineicosane	296 [M ⁺] (3.0%), 253 [M-C ₃ H ₇] ⁺ (0.7%), 239 [M-C ₄ H ₉] ⁺ (1.3%), 225 [M-C ₅ H ₁₁] ⁺ (1.0%), 211 [M-C ₆ H ₁₃] ⁺ (1.3%), 197 [M-C ₇ H ₁₅] ⁺ (2.3%), 183 [M-C ₈ H ₁₇] ⁺ (3.0%), 169 [M-C ₉ H ₁₉] ⁺ (3.7%), 155 [M-C ₁₀ H ₂₁] ⁺ (4.7%), 141 [M-C ₁₁ H ₂₃] ⁺ (6.7%), 127 [M-C ₁₂ H ₂₅] ⁺ (9.3%), 113 [M-C ₁₃ H ₂₇] ⁺ (13.3%), 99 [M-C ₁₄ H ₂₉] ⁺ (24.0%), 85 [M-C ₁₅ H ₃₁] ⁺ (65.7%), 71 [M-C ₁₆ H ₃₃] ⁺ (83.3%), 57 [M-C ₁₇ H ₃₅] ⁺ (100%)
13 n-docosane	310 [M ⁺] (1.3%), 288 (0.3%), 253 [M-C ₄ H ₉] ⁺ (0.7%), 225 [M-C ₆ H ₁₃] ⁺ (0.7%), 211 [M-C ₇ H ₁₅] ⁺ (0.7%), 183 [M-C ₉ H ₁₉] ⁺ (3.3%), 155 [M-C ₁₁ H ₂₃] ⁺ (4.7%), 127 [M-C ₁₃ H ₂₇] ⁺ (10.0%), 113 [M-C ₁₄ H ₂₉] ⁺ (13.3%), 99 [M-C ₁₅ H ₃₁] ⁺ (21.7%), 85 [M-C ₁₆ H ₃₃] ⁺ (83.3%), 71 [M-C ₁₇ H ₃₅] ⁺ (83.0%), 57 [M-C ₁₈ H ₃₇] ⁺ (100%)

14	n-tricosane	324 [M ⁺] (2.7%), 281 [M-C ₃ H ₇] ⁺ (0.3%), 253 [M-C ₅ H ₁₁] ⁺ (0.3%), 225 [M-C ₇ H ₁₅] ⁺ (0.3%), 197 [M-C ₉ H ₁₉] ⁺ (2.7%), 169 [M-C ₁₁ H ₂₃] ⁺ (3.7%), 155 [M-C ₁₂ H ₂₅] ⁺ (6.0%), 141 [M-C ₁₃ H ₂₇] ⁺ (6.7%), 127 [M-C ₁₄ H ₂₉] ⁺ (10.0%), 113 [M-C ₁₅ H ₃₁] ⁺ (13.7%), 99 [M-C ₁₆ H ₃₃] ⁺ (23.0%), 85 [M-C ₁₇ H ₃₅] ⁺ (70.0%), 71 [M-C ₁₈ H ₃₇] ⁺ (83.3%), 57 [M-C ₁₉ H ₃₉] ⁺ (100%)
15	n-tetracosane	338 [M ⁺] (1.3%), 281 [M-C ₄ H ₉] ⁺ (0.7%), 253 [M-C ₆ H ₁₃] ⁺ (0.7%), 225 [M-C ₈ H ₁₇] ⁺ (0.7%), 197 [M-C ₁₀ H ₂₁] ⁺ (3.0%), 169 [M-C ₁₂ H ₂₅] ⁺ (3.7%), 155 [M-C ₁₃ H ₂₇] ⁺ (6.3%), 141 [M-C ₁₄ H ₂₉] ⁺ (7.0%), 127 [M-C ₁₅ H ₃₁] ⁺ (10.3%), 113 [M-C ₁₆ H ₃₃] ⁺ (14.0%), 99 [M-C ₁₇ H ₃₅] ⁺ (23.0%), 85 [M-C ₁₈ H ₃₇] ⁺ (70.0%), 71 [M-C ₁₉ H ₃₉] ⁺ (83.3%), 57 [M-C ₂₀ H ₄₁] ⁺ (100%)
16	n-pentacosane	352 [M ⁺] (1.0%), 323 [M-C ₂ H ₅] ⁺ (0.3%), 281 [M-C ₅ H ₁₁] ⁺ (0.7%), 253 [M-C ₇ H ₁₅] ⁺ (0.7%), 225 [M-C ₉ H ₁₉] ⁺ (0.7%), 197 [M-C ₁₁ H ₂₃] ⁺ (3.0%), 169 [M-C ₁₃ H ₂₇] ⁺ (3.7%), 155 [M-C ₁₄ H ₂₉] ⁺ (6.3%), 141 [M-C ₁₅ H ₃₁] ⁺ (7.0%), 127 [M-C ₁₆ H ₃₃] ⁺ (10.3%), 113 [M-C ₁₇ H ₃₅] ⁺ (14.0%), 99 [M-C ₁₈ H ₃₇] ⁺ (23.0%), 85 [M-C ₁₉ H ₃₉] ⁺ (70.0%), 71 [M-C ₂₀ H ₄₁] ⁺ (83.3%), 57 [M-C ₂₁ H ₄₃] ⁺ (100%)
17	n-hexacosane	366 [M ⁺] (0.7%), 309 [M-C ₃ H ₇] ⁺ (0.3%), 281 [M-C ₆ H ₁₃] ⁺ (2.7%), 253 [M-C ₇ H ₁₅] ⁺ (2.7%), 225 [M-C ₁₀ H ₂₁] ⁺ (2.7%), 197 [M-C ₁₂ H ₂₅] ⁺ (3.0%), 169 [M-C ₁₄ H ₂₉] ⁺ (4.0%), 141 [M-C ₁₆ H ₃₃] ⁺ (7.0%), 127 [M-C ₁₇ H ₃₅] ⁺ (10.3%), 113 [M-C ₁₈ H ₃₇] ⁺ (14.0%), 99 [M-C ₁₉ H ₃₉] ⁺ (23.3%), 85 [M-C ₂₀ H ₄₁] ⁺ (66.7%), 71 [M-C ₂₁ H ₄₃] ⁺ (83.3%), 57 [M-C ₂₂ H ₄₅] ⁺ (100%)
18	n-heptacosane	380 [M ⁺] (0.7%), 337 [M-C ₃ H ₇] ⁺ (0.3%), 323 [M-C ₄ H ₉] ⁺ (0.3%), 295 [M-C ₆ H ₁₃] ⁺ (3.0%), 281 [M-C ₇ H ₁₅] ⁺ (3.0%), 267 [M-C ₈ H ₁₇] ⁺ (0.4%), 225 [M-C ₁₁ H ₂₃] ⁺ (1.3%), 211 [M-C ₁₂ H ₂₅] ⁺ (3.3%), 197 [M-C ₁₃ H ₂₇] ⁺ (3.3%), 169 [M-C ₁₅ H ₃₁] ⁺ (4.3%), 141 [M-C ₁₇ H ₃₅] ⁺ (7.0%), 113 [M-C ₁₉ H ₃₉] ⁺ (14.7%), 99 [M-C ₂₀ H ₄₁] ⁺ (23.3%), 85 [M-C ₂₁ H ₄₃] ⁺ (64.7%), 71 [M-C ₂₂ H ₄₅] ⁺ (82.7%), 57 [M-C ₂₃ H ₄₇] ⁺ (100%)
19	n-octacosane	394 [M ⁺] (0.3%), 355 (0.3%), 309 [M-C ₆ H ₁₃] ⁺ (0.7%), 281 [M-C ₈ H ₁₇] ⁺ (6.3%), 253 [M-C ₁₀ H ₂₁] ⁺ (3.0%), 225 [M-C ₁₂ H ₂₅] ⁺ (2.7%), 207 [M-C ₁₄ H ₁₉] ⁺ (13.0%), 169 [M-C ₁₆ H ₃₃] ⁺ (4.3%), 141 [M-C ₁₈ H ₃₇] ⁺ (7.0%), 127 [M-C ₁₉ H ₃₉] ⁺ (10.3%), 113 [M-C ₂₀ H ₄₁] ⁺ (14.3%), 99 [M-C ₂₁ H ₄₃] ⁺ (23.3%), 85 [M-C ₂₂ H ₄₅] ⁺ (63.3%), 71 [M-C ₂₃ H ₄₇] ⁺ (80.3%), 57 [M-C ₂₄ H ₄₉] ⁺ (100%)
20	n-nonacosane	408 [M ⁺] (0.3%), 355 (0.7%), 309 [M-C ₇ H ₁₅] ⁺ (0.7%), 281 [M-C ₉ H ₁₉] ⁺ (6.3%), 253 [M-C ₁₁ H ₂₃] ⁺ (3.0%), 225 [M-C ₁₃ H ₂₇] ⁺ (0.7%), 207 [M-C ₁₅ H ₂₁] ⁺ (10.7%), 169 [M-C ₁₇ H ₃₅] ⁺ (3.7%), 141 [M-C ₁₉ H ₃₉] ⁺ (8.3%), 127 [M-C ₂₀ H ₄₁] ⁺ (11.0%), 113 [M-C ₂₁ H ₄₃] ⁺ (16.0%), 99 [M-C ₂₂ H ₄₅] ⁺ (23.3%), 85 [M-C ₂₃ H ₄₇] ⁺ (64.3%), 71 [M-C ₂₄ H ₄₉] ⁺ (83.3%), 57 [M-C ₂₅ H ₅₁] ⁺ (100%)
21	2-methyldodecane	184 [M ⁺] (0.3%), 169 [M-CH ₃] ⁺ (6.7%), 141 [M-C ₃ H ₇] ⁺ (13.5%), 127 [M-C ₄ H ₉] ⁺ (3.3%), 113 [M-C ₅ H ₁₁] ⁺ (8.3%), 99 [M-C ₆ H ₁₃] ⁺ (19.7%), 85 [M-C ₇ H ₁₅] ⁺ (60.0%), 71 [M-C ₈ H ₁₇] ⁺ (76.7%), 57 [M-C ₉ H ₁₉] ⁺ (100%)
22	2-methylhexadecane	240 [M ⁺] (0.2%), 225 [M-CH ₃] ⁺ (4.7%), 197 [M-C ₃ H ₇] ⁺ (13.3%), 184 [M-C ₄ H ₈] ⁺ (10.0%), 169 [M-C ₅ H ₁₁] ⁺ (8.3%), 155 [M-C ₆ H ₁₃] ⁺ (13.3%), 141 [M-C ₇ H ₁₅] ⁺ (10.0%), 127 [M-C ₈ H ₁₇] ⁺ (13.3%), 113 [M-C ₉ H ₁₉] ⁺ (16.7%), 99 [M-C ₁₀ H ₂₁] ⁺ (22.7%), 85 [M-C ₁₁ H ₂₃] ⁺ (54.3%), 71 [M-C ₁₂ H ₂₅] ⁺ (64.3%), 57 [M-C ₁₃ H ₂₇] ⁺ (100%)
23	3-methyloctadecane	268 [M ⁺] (0.3%), 253 [M-CH ₃] ⁺ (3.3%), 239 [M-C ₂ H ₅] ⁺ (1.0%), 211 [M-C ₄ H ₉] ⁺ (3.3%), 197 [M-C ₅ H ₁₁] ⁺ (4.0%), 183 [M-C ₆ H ₁₃] ⁺ (7.7%), 169 [M-C ₇ H ₁₅] ⁺ (11.3%), 155 [M-C ₈ H ₁₇] ⁺ (10.0%), 141 [M-C ₉ H ₁₉] ⁺ (10.7%), 127 [M-C ₁₀ H ₂₁] ⁺ (11.0%), 113 [M-C ₁₁ H ₂₃] ⁺ (13.3%), 99 [M-C ₁₂ H ₂₅] ⁺ (21.7%), 85 [M-C ₁₃ H ₂₇] ⁺ (56.0%), 71 [M-C ₁₄ H ₂₉] ⁺ (70.7%), 57 [M-C ₁₅ H ₃₁] ⁺ (100%)
24	2,6-dimethyldodecane	198 [M ⁺] (0.75%), 183 [M-CH ₃] ⁺ (0.2%), 155 [M-C ₃ H ₇] ⁺ (1.3%), 141 [M-C ₄ H ₉] ⁺ (1.0%), 127 [M-C ₅ H ₁₁] ⁺ (2.7%), 113 [M-C ₆ H ₁₃] ⁺ (16.7%), 85 [M-C ₇ H ₁₅] ⁺ (14.3%), 71 [M-C ₉ H ₁₉] ⁺ (92.7%), 57 (100%)
25	n-hexadecanoic acid	256 [M ⁺] (29.3%), 241 [M-CH ₃] ⁺ (0.7%), 227 [M-C ₂ H ₅] ⁺ (9.0%), 213 [M-C ₃ H ₇] ⁺ (10.5%), 199 [M-C ₄ H ₇] ⁺ (10.0%), 185 [M-C ₅ H ₁₁] ⁺ (21.3%), 171 [M-C ₆ H ₁₃] ⁺ (21.0%), 157 [M-C ₇ H ₁₅] ⁺ (23.3%), 143 [M-C ₈ H ₁₇] ⁺ (11.3%), 139 [M-C ₄ H ₉] ⁺ (3.3%), 129 [M-C ₉ H ₁₉] ⁺ (56.7%), 115 [M-C ₁₀ H ₂₁] ⁺ (22.7%), 97 [M-C ₇ H ₁₅] ⁺ (37.0%), 85 [M-C ₈ H ₁₅] ⁺ (39.3%), 83 [M-C ₈ H ₁₇] ⁺ (46.7%), 73 [M-C ₁₃ H ₂₇] ⁺ (100%), 57 [M-C ₁₀ H ₁₉] ⁺ (82.0%), 55 [M-C ₁₀ H ₂₁] ⁺ (66.7%)
26	9-octadecenoic acid	282 [M ⁺] (0.3%), 265 [M-OH] ⁺ (3.3%), 223 [M-C ₂ H ₃ O ₂] ⁺ (0.3%), 181 [M-C ₅ H ₉ O ₂] ⁺ (0.3%), 153 [M-C ₇ H ₁₃ O ₂] ⁺ (3.3%), 99 [M-C ₁₁ H ₁₉ O ₂] ⁺ (4.0%), 85 [M-C ₁₂ H ₂₁ O ₂] ⁺ (13.35), 55 (100%), 43 [M-C ₁₅ H ₂₇ O ₂] ⁺ (53.3%)
27	methyl-14-methylpentadecanoate	270 [M ⁺] (40.6%), 241 (13.1%), 227 (29.8%), 199 [M-C ₃ H ₅ O ₂] ⁺ (12.0%), 187 (13.5%), 149 (20.1%), 135 (31.6%), 129 [M-C ₁₀ H ₂₁] ⁺ (24.8%), 87 [M-C ₁₃ H ₂₇] ⁺ (100%), 69 (83.5), 57 [M-C ₁₃ H ₂₅ O ₂] ⁺ (38.4%)
28	methylpalmitate	270 [M ⁺] (13.3%), 239 [M-CH ₃ O] ⁺ (4.73%), 227 [M-C ₃ H ₇] ⁺ (10.0%), 213 [M-C ₄ H ₉] ⁺ (3.0%), 199 [M-C ₅ H ₁₁] ⁺ (6.7%), 185 [M-C ₆ H ₁₃] ⁺ (6.7%), 171 [M-C ₇ H ₁₅] ⁺ (6.7%), 157 [M-C ₈ H ₁₇] ⁺ (3.3%), 143

		[M-C ₉ H ₁₉] ⁺ (18.0%), 129 [M-C ₁₀ H ₂₁] ⁺ (6.7%), 115 [M-C ₁₁ H ₂₃] ⁺ (3.3%), 97 [M-C ₁₂ H ₂₅] ⁺ (6.7%), 87 [M-C ₁₃ H ₂₇] ⁺ (70.0%), 74 (100%), 55 (30.0%), 43 (40.0%), 29 (3.3%)
29	methyloctadeca-10,13-diynoate	290 [M ⁺] (0.2%), 284 (0.2%), 231 [M-C ₂ H ₃ O ₂] ⁺ (2.6%), 203 [M-C ₄ H ₇ O ₂] ⁺ (6.4%), 147 [M-C ₈ H ₁₅ O ₂] ⁺ (7.7%), 133 [M-C ₉ H ₁₇ O ₂] ⁺ (35.4%), 119 [M-C ₁₀ H ₁₉ O ₂] ⁺ (14.3%), 105 [M-C ₁₁ H ₂₁ O ₂] ⁺ (68.0%), 91 (100%), 69 [M-C ₁₄ H ₂₁ O ₂] ⁺ (11.0%), 57 [M-C ₁₅ H ₂₁ O ₂] ⁺ (17.6%)
30	methyloctadeca-12,15-diynoate	290 [M ⁺] (0.2%), 259 [M-CH ₃ O] ⁺ (0.2%), 232 (6.1%), 203 [M-C ₄ H ₇ O ₂] ⁺ (0.4%), 175 [M-C ₆ H ₁₁ O ₂] ⁺ (4.6%), 147 [M-C ₈ H ₁₅ O ₂] ⁺ (10.3%), 119 [M-C ₁₀ H ₁₉ O ₂] ⁺ (16.2%), 105 [M-C ₁₁ H ₂₁ O ₂] ⁺ (24.6%), 91 [M-C ₁₂ H ₁₃ O ₂] ⁺ (100%), 69 (11.0%), 57 (16.2%)
31	methyllinolenate	292 [M ⁺] (0.3%), 277 [M-CH ₃] ⁺ (0.3%), 261 [M-CH ₃ O] ⁺ (3.3%), 191 [M-C ₃ H ₉ O ₂] ⁺ (3.3%), 163 [M-C ₇ H ₁₃ O ₂] ⁺ (6.3%), 149 [M-C ₈ H ₁₅ O ₂] ⁺ (16.7%), 135 [M-C ₉ H ₁₇ O ₂] ⁺ (13.7%), 109 [M-C ₁₁ H ₁₉ O ₂] ⁺ (16.7%), 95 [M-C ₁₂ H ₂₁ O ₂] ⁺ (60.0%), 79 (100%), 55 [M-C ₁₅ H ₂₅ O ₂] ⁺ (46.7%), 29 [M-C ₁₇ H ₂₇ O ₂] ⁺ (16.7%)
32	methylstearate	298 [M ⁺] (9.1%), 267 [M-CH ₃ O] ⁺ (2.2%), 255 [M-C ₃ H ₇] ⁺ (6.5%), 199 [M-C ₇ H ₁₅] ⁺ (7.7%), 165 (6.8%), 143 [M-C ₁₁ H ₂₃] ⁺ (22.6%), 129 (10.8%), 101 [M-C ₁₄ H ₂₉] ⁺ (9.1%), 97 (19.0%), 87 [M-C ₁₅ H ₃₁] ⁺ (70.4%), 74 (100%), 57 [M-C ₁₅ H ₂₉ O ₂] ⁺ (36.1%)
33	dehydroxyloliolide	180 [M ⁺] (16.7%), 165 [M-CH ₃] ⁺ (6.7%), 155 (23.3%), 147 (26.7%), 135 (37.3%), 124 (16.7%), 111 (91.7%), 97 (36.7%), 83 (100%), 69 (56%), 55 (70.0%)
34	geranylacetone	194 [M ⁺] (0.3%), 179 [M-CH ₃] ⁺ (0.7%), 151 [M-C ₂ H ₃ O] ⁺ (0.7%), 137 [M-C ₃ H ₅ O] ⁺ (10.0%), 123 [M-C ₄ H ₇ O] ⁺ (8.0%), 83 [M-C ₇ H ₁₁ O] ⁺ (9.7%), 69 [M-C ₈ H ₁₃ O] ⁺ (100%), 43 (60.0%)
35	loliolide	196 [M ⁺] (10.3%), 179 [M-OH] ⁺ (9.0%), 163 (30.0%), 153 [M-C ₃ H ₇] ⁺ (20.7%), 140 (36.0%), 125 (16.7%), 111 (100%), 97 (37.3%), 85 (40.0%), 67 (36.7%), 57 (44.0%)
36	trans-β-caryophyllene	204 [M ⁺] (0.5%), 189 [M-CH ₃] ⁺ (11.3%), 175 [M-C ₂ H ₅] ⁺ (18.5%), 147 [M-C ₄ H ₉] ⁺ (54.2%), 135 [M-C ₅ H ₉] ⁺ (27.1%), 120 [M-C ₆ H ₁₂] ⁺ (85.8%), 93 [M-C ₈ H ₁₅] ⁺ (32.3%), 79 [M-C ₉ H ₁₇] ⁺ (100%), 67 [M-C ₁₀ H ₁₇] ⁺ (53.3%), 53 [M-C ₁₁ H ₁₉] ⁺ (29.3%)
37	4(15),6-cadinadiene	204 [M ⁺] (25.28%), 183 (3.4%), 161 [M-C ₃ H ₇] ⁺ (100%), 133 (15.8%), 119 [M-C ₆ H ₁₃] ⁺ (11.3%), 105 [M-C ₇ H ₁₅] ⁺ (68.0%), 93 [M-C ₈ H ₁₅] ⁺ (40.6%), 71 [M-C ₁₀ H ₁₆] ⁺ (26.4%), 57 (32.7%)
38	2,6,10-trimethyldodecane	212 [M ⁺] (3.3%), 197 [M-CH ₃] ⁺ (1.0%), 183 [M-C ₂ H ₅] ⁺ (4.0%), 169 [M-C ₃ H ₇] ⁺ (0.1%), 155 [M-C ₄ H ₉] ⁺ (3.3%), 141 [M-C ₅ H ₁₁] ⁺ (3.3%), 127 [M-C ₆ H ₁₃] ⁺ (16.7%), 113 [M-C ₇ H ₁₅] ⁺ (11.7%), 97 [M-C ₈ H ₁₇] ⁺ (21.0%), 85 [M-C ₉ H ₁₉] ⁺ (32.3%), 71 [M-C ₁₀ H ₂₁] ⁺ (73.3%), 57 [M-C ₁₁ H ₂₃] ⁺ (100%), 43 [M-C ₁₂ H ₂₄] ⁺ (83.3%)
39	capnellane-5-α-ol	222 [M ⁺] (17.3%), 207 [M-CH ₃] ⁺ (89.3%), 189 (16.7%), 151 [M-C ₅ H ₁₁] ⁺ (23.6%), 135 (23.3%), 109 [M-C ₈ H ₁₇] ⁺ (66.0%), 95 [M-C ₉ H ₁₉] ⁺ (84.0%), 81 [M-C ₁₀ H ₂₁] ⁺ (100%), 55 [M-C ₁₂ H ₂₃] ⁺ (59.35)
40	pristine	268 [M ⁺] (0.2%), 232 (0.2%), 197 [M-C ₅ H ₁₁] ⁺ (0.3%), 183 [M-C ₆ H ₁₃] ⁺ (16.7%), 169 [M-C ₇ H ₁₅] ⁺ (3.7%), 155 [M-C ₈ H ₁₇] ⁺ (6.3%), 141 [M-C ₉ H ₁₉] ⁺ (6.0%), 127 [M-C ₁₀ H ₂₁] ⁺ (10.0%), 113 [M-C ₁₁ H ₂₃] ⁺ (20.0%), 99 [M-C ₁₂ H ₂₅] ⁺ (16.7%), 85 [M-C ₁₃ H ₂₇] ⁺ (40.7%), 71 [M-C ₁₄ H ₂₉] ⁺ (93.0%), 57 [M-C ₁₅ H ₃₁] ⁺ (100%)
41	6,10,14-trimethyl-2-pentadecanone	268 [M ⁺] (2.9%), 253 [M-CH ₃] ⁺ (2.3%), 211 [M-C ₄ H ₉] ⁺ (2.7%), 155 [M-C ₈ H ₁₇] ⁺ (6.8%), 124 (38.4%), 109 (63.2%), 95 (31.6%), 85 [M-C ₁₃ H ₂₇] ⁺ (41.1%), 71 [M-C ₁₄ H ₂₉] ⁺ (67.7%), 57 [M-C ₁₅ H ₃₁] ⁺ (59.8%)
42	neophytadiene	278 [M ⁺] (2.7%), 263 [M-CH ₃] ⁺ (0.7%), 208 [M-C ₅ H ₁₀] ⁺ (1.0%), 193 [M-C ₆ H ₁₃] ⁺ (3.0%), 179 [M-C ₇ H ₁₅] ⁺ (4.7%), 165 [M-C ₈ H ₁₇] ⁺ (3.3%), 151 [M-C ₉ H ₁₉] ⁺ (3.3%), 137 [M-C ₁₀ H ₂₁] ⁺ (13.0%), 123 [M-C ₁₁ H ₂₃] ⁺ (66.7%), 109 [M-C ₁₂ H ₂₅] ⁺ (33.3%), 95 [M-C ₁₃ H ₂₇] ⁺ (96.7%), 82 [M-C ₁₄ H ₂₈] ⁺ (83.3%), 68 [M-C ₁₅ H ₃₀] ⁺ (100%), 55 [M-C ₁₆ H ₃₁] ⁺ (66.7%)
43	phytane	282 [M ⁺] (0.3%), 197 (1.3%), 183 [M-C ₇ H ₁₅] ⁺ (3.3%), 169 [M-C ₈ H ₁₇] ⁺ (6.3%), 141 [M-C ₁₀ H ₂₁] ⁺ (10.0%), 127 [M-C ₁₁ H ₂₃] ⁺ (6.7%), 113 (15.7%), 99 [M-C ₁₃ H ₂₇] ⁺ (20.0%), 85 [M-C ₁₄ H ₂₉] ⁺ (54.0%), 71 [M-C ₁₅ H ₃₁] ⁺ (87.7%), 57 [M-C ₁₆ H ₃₃] ⁺ (100%)
44	crocetane	282 [M ⁺] (0.2%), 232 (3.0%), 208 (3.3), 183 [M-C ₇ H ₁₅] ⁺ (6.7%), 169 [M-C ₈ H ₁₇] ⁺ (10.7%), 155 [M-C ₉ H ₁₉] ⁺ (6.7%), 133 [M-C ₁₁ H ₁₇] ⁺ (17.7%), 91 (40.0%), 71 [M-C ₁₅ H ₃₁] ⁺ (60.0%), 57 [M-C ₁₆ H ₃₃] ⁺ (100%)
45	14-β-H-pregnane	288 [M ⁺] (0.3%), 280 (5.0%), 250 (11.7%), 220 (30.7%), 189 (20.0%), 165 (20.7%), 141 (17.0%), 111 (43.0%), 83 (75.0%), 57 (100%)
46	phytol	296 [M ⁺] (0.3%), 279 [M-OH] ⁺ (0.3%), 211 [M-C ₅ H ₉ O] ⁺ (0.3%), 197 [M-C ₆ H ₁₁ O] ⁺ (2.0%), 123 [M-C ₁₂ H ₁₃ O] ⁺ (22.3%), 111 [M-C ₁₂ H ₂₅ O] ⁺ (13.3%), 95 [M-C ₁₄ H ₁₇ O] ⁺ (16.7%), 83 [M-C ₁₅ H ₁₇ O] ⁺ (24.0%), 71 [M-C ₁₅ H ₂₉ O] ⁺ (100%), 57 [M-C ₁₆ H ₃₁ O] ⁺ (33.3%), 43 [M-C ₁₇ H ₃₃ O] ⁺ (29.0%)

47	4,8,12,16-tetramethyl heptadecan-1,4-olide	324 [M ⁺] (0.3%), 225 [M-C ₅ H ₇ O] ⁺ (0.3%), 197 [M-C ₇ H ₁₁ O ₂] ⁺ (0.7%), 155 [M-C ₁₀ H ₁₇ O ₂] ⁺ (2.3%), 127 [M-C ₁₂ H ₂₁ O ₂] ⁺ (6.7%), 113 [M-C ₅ H ₇ O] ⁺ (3.3%), 99 (100%), 85 [M-C ₁₅ H ₂₇ O ₂] ⁺ (16.0%), 57 [M-C ₁₇ H ₃₁ O ₂] ⁺ (30.0%)
48	2-[4-methyl-6-(2,6,6-trimethylcyclohex-1-enyl)hexa-1,3,5-trienyl]cyclohex-1-en-1-carboxaldehyde	324 [M ⁺] (0.2%), 274 (3.2%), 243 (2.3%), 217 (6.1%), 192 (3.2%), 165 (6.8%), 147 (33.9%), 133 (9.1%), 119 (18.5%), 105 (31.6%), 91 (100%), 69 (17.4%), 55 (24.8%)
49	gibberline A3	346 [M ⁺] (1.0%), 327 (8.0%), 300 (2.0%), 281 (46.7%), 253 (13.3%), 207 (100%), 177 (4.0%), 159 (6.7%), 96 (10.0%), 73 (18.3%), 55 (11.0%)
50	Squalene	410 [M ⁺] (0.7%), 367 [M-C ₃ H ₇] (0.3%), 341 [M-C ₅ H ₉] (2.7%), 281 (3.3%), 207 [M-C ₁₅ H ₂₅] (6.7%), 191 [M-C ₁₆ H ₂₇] (3.3%), 137 [M-C ₂₀ H ₃₃] (16.3%), 123 [M-C ₂₁ H ₃₅] (10.0%), 69 [M-C ₂₅ H ₄₁] (100%), 55 [M-C ₂₆ H ₄₃] (2.3%)
51	vitamin E	430 [M ⁺] (71.0%), 405 [M-C ₂ H ₆] ⁺ (0.3%), 355 (0.7%), 327 (0.3%), 281 [M-C ₁₀ H ₁₂ O] ⁺ (6.3%), 252 [M-C ₁₁ H ₁₄ O ₂] ⁺ (0.3%), 225 [M-C ₁₃ H ₁₇ O ₂] ⁺ (0.3%), 205 [C ₁₃ H ₁₇ O ₂] ⁺ (10.0%), 191 (6.7%), 165 [C ₁₀ H ₁₃ O ₂] ⁺ (100%), 135 (6.7%), 71 [M-C ₂₄ H ₃₉ O ₂] ⁺ (3.0%), 57 [M-C ₂₅ H ₄₁ O ₂] ⁺ (6.7%)
52	3,4,3',4'tetrahydrospirilloxanthin	600 [M ⁺] (1.8%), 582 (5.3%), 569 [M-CH ₃ O] ⁺ (2.2%), 499 [M-C ₆ H ₁₃ O] ⁺ (2.2%), 470 [M-C ₈ H ₁₈ O] ⁺ (6.8%), 459 [M-C ₉ H ₁₇ O] ⁺ (6.3%), 429 (43.3%), 352 [M-C ₁₇ H ₂₈ O] ⁺ (11.5%), 340 [M-C ₁₉ H ₃₀ O] ⁺ (33.9%), 326 (27.1%), 282 (16.2%), 259 (27.1%), 241 (20.3%), 215 (49.2%), 193 (33.3%), 169 [M-C ₃₁ H ₄₃ O] ⁺ (29.3%), 121 (61.0%), 115 [M-C ₃₅ H ₄₉ O] ⁺ (42.7%), 105 (38.3%), 69 (100%)
53	1-methyl-9H-fluorine	180 [M ⁺] (70.0%), 165 [M-CH ₃] ⁺ (100%), 152 [M-C ₂ H ₄] ⁺ (15.7%), 141 (7.3%), 128 (10.0%), 111 (10.0%), 97 (26.0%), 83 (20.0%), 69 (20.0%), 57 (23.3%)
54	Diocyladipate	341 [M ⁺] (4.5%), 283 (2.5%), 241 [M-C ₈ H ₁₇ O] ⁺ (18.1%), 199 [M-C ₁₀ H ₁₉ O ₂] ⁺ (9.7%), 167 (5.6%), 147 (40.6%), 129 [M-C ₁₄ H ₂₅ O ₃] ⁺ (100%), 113 [M-C ₁₄ H ₂₅ O ₄] ⁺ (2.2%), 83 (32.5), 71 [M-C ₁₀ H ₁₉ O ₂] ⁺ (36.1%), 57 [M-C ₁₇ H ₃₁ O ₄] ⁺ (56.4%)

3.4.2 Antioxidant activity assessment

3.4.2.1 Free radical scavenging method (ABTS)

For each of the investigated fractions (2 µl) of ABTS solution was added to 3 µl MnO₂ solution (25 mg/ µl), all were prepared in (5 µl) aqueous phosphate buffer solution (pH 7, 0.1 M). The mixture was shaken, centrifuged, filtered and the absorbance of the resulting green blue solution (ABTS radical solution) at 734 nm was adjusted to approx. ca. 0.5. Then, 50 µl of (2 µM) solution of the tested fraction and the isolated compound in spectroscopic grade MeOH / phosphate buffer (1:1) was added. The absorbance was measured and the reduction in color intensity was expressed as inhibition percentage. L-ascorbic acid was used as standard antioxidant (positive control). Blank sample was run without ABTS and using MeOH / phosphate buffer (1:1) instead of tested fractions. Negative control was run with ABTS and MeOH / phosphate buffer (1:1) only.

3.4.3 Cytotoxicity assay

The cell lines MCF-7, HePG-2 and PC3 were used to determine the inhibitory effects of extracts on cell growth using the MTT assay (Mosmann and Immunol, 1983; Denizot et al., 1986). This colorimetric assay is based on the conversion of the yellow tetrazolium bromide (MTT) to a purple formazan derivative by mitochondrial succinate dehydrogenase in viable cells. Cell lines were cultured in RPMI-1640 medium with 10% fetal bovine serum. Antibiotics added were 100 units / µl penicillin

and 100 µg/µl streptomycin at 37°C in a 5% CO₂ incubator. The cell lines were seeded in a 96-well plate at a density of a 1.0x10⁴ cells/well at 37 °C for 48 h under 5% CO₂. After incubation the cells were treated with different concentration of the fractions and isolated compound and incubated for 24 h. After 24 h of the treatment, 20 µl of MTT solution at 5 mg/µl was added and incubated for 4 h. Dimethyl sulfoxide (DMSO) in volume of 100µl is added into each well to dissolve the purple formazan formed. The colorimetric assay is measured and recorded at absorbance of 570 nm using a plate reader (EXL 800). The relative cell viability was calculated from the formula:

$$\% \text{ Relative cell viability} = \frac{\text{Absorbance 570 of treated samples}}{\text{Absorbance 570 of untreated sample}} \times 100$$

4. REFERENCES

- [1] Angiosperm Phylogeny Website. Retrieved 15 July 2014.
- [2] Fryxell P. A. 1997a. The American genera of MalvaceaeII. Brittonia. 49: 204-269.
- [3] Al-Gohary H. 2008. Floristic Composition of Eleven Wadis in Gebel Elba, Egypt. IJAB. 10: 151-160.
- [4] Shaltout K, Sheded G. M., and Salem A. 2010. Vegetation spatial heterogeneity in a hyper arid Biosphere Reserve area in North Africa. Acta Bot. Croat. 69: 31-46.

- [5] Alshamy M. M. 2016. Ecological study on wild vegetation of palm fields across some oases in western Desert, Egypt. *CATRINA*. 15: in press.
- [6] Qing ma shu. 2007. *Flora of China*. 12: 275-279.
- [7] Boulos L. 2002. *Flora of Egypt*. Al Hadara Publishing, Cairo, Egypt. 3: 229-231.
- [8] Vaghasiya Y., Chanda S., and Turk J. 2007. *Biol.* 31: 243-248.
- [9] Ali S., Yasmeen S., Afza N., Malik A., Iqbal L., Lateef M., Riaz N., and Ashraf M. J. 2009. *Asian Nat Prod Res.* 11: 457-464.
- [10] Eldahshan O. A. 2010. Isolation and Structure Elucidation of Phenolic Compounds of Carob Leaves Grown in Egypt. *Curr. Res. J. Biol. Sci.* 3: ISSN: 2041-0778.
- [11] Ibrahim A., Khalifa S., khafagi I., Youssef D., Khan S., Mesbah M. k., and Khan I. 2008. Microbial Metabolism of Biologically Active Secondary Metabolites from Nerium oleander L. *Chem. Pharm. Bull.* 56: 1256-1257.
- [12] Lissi E., Modak B., Torres R., Escobar J., and Urza A. 1999. *Free Radical Res.* 30: 471 – 477.
- [13] El-Gazar A. B. A., Youssef A. M. S., Youssef M. M., Abu-Hashem A. A., Badria F. A., and Eur. J. 2009. *Med. Chem.* 44: 609 – 624.
- [14] Aeschlach R., Loliger J., Scott C. B., Murcia A., Butler J., Halliwell B., and Aruoma I. O. 1994. *Food Chem. Toxicol.* 32: 31 – 36.
- [15] Stylianakis A., Kolocouris N., Kolocouris G., Fytas G. B., Foscolos E., Padalko J., Neyts D., and Clerq E. 2003. *Bioorg. Med. Chem. Lett.* 13: 1699 – 1703.
- [16] Mosmann T., and Immunol J. 1983. *Methods.* 65: 55-63.
- [17] Denizot F., Lang R., and Immunol J. 1986. *Methods.* 22: 271-277.
- [18] Mauceri H. J., Hanna N. N., Beckett M. A., Gorski D. H., Staba M. J., Stellato K. A., Bigelow K., Heimann R, Gately S., Dhanabal M., Soff G. A., Sukhatme V. P., Kufe D. W., and Weichselbaum R. R. 1998. Combined effects of agiostatin and ionizing radiation in antitumour therapy. *Nature.* 394: 287-291.

Implementation of Advanced Encryption Standard Algorithm with Key Length of 256 Bits for Preventing Data Loss in an Organization

Owusu Nyarko- Boateng
Innerjoy Digital Systems
Sunyani, Ghana

Michael Asante
Department of Computer
Science
Kwame Nkrumah University of
Science and Technology,
Kumasi, Ghana

Isaac Kofi Nti
Department of
Electrical/Electronic
Engineering
Sunyani Technical University
Sunyani, Ghana

Abstract: Data and network security is one among the foremost necessary factors in today's business world. In recent, businesses and firms like financial institutions, law firms, schools, health sectors, telecommunications, mining and a number of government agencies want a strategic security technique of managing its data. Organizations managing bigger monetary information, bio-data and alternative relevant info are losing its valuable information or data at rest, in usage or in motion to unauthorized parties or competitors as a result of activities of hackers. Organizations are losing millions of dollars as a result of unprotected data that gets into the hands of malicious persons [1]. Data protection in an organization has become vital in today's business. In order to possess secure information, this information should be protected in order that although malicious persons get access to the info, it becomes wealth-less and useless to them. Advanced Encryption Standard (AES), could be a scientific discipline rule that may be used for secured data and communication in an organization, it uses same key that's isobilateral key for transmission additionally as reception. The AES rule is capable of using cryptographic keys of 128, 192, and 256 bits, this paper implement AES block cipher of 256-bits and 256-bit key size, developed with C# as a front-end client machine and MS SQL used for the database as a back-end machine.

Keywords: Cryptography; AES-algorithm; Decryption; Encryption; Cipher; 256-bit-key

1. INTRODUCTION

The scrambling of plaintext delivers a safe and nice significance for secured data and communication. Use of scientific discipline algorithms is completed for the aim of security in varied applications like secured optical disk content, ATM, etc. Secured data and communication in an organization is one in all the foremost necessary things in present day business and its requirement is rapidly increasing [2] [3]. With the introduction of LAN, WAN, MAN and internet technology in recent years, the computer network communication is exposed to unwanted people giving them access to pose different kinds of attacks on personal and organizational data in a network environment [4]. Every individual desires, their information to be secured and privacy should be maintained. This demand is consummated by the employment of cryptography. Numerous security systems are needed to guard a shared information in an organization. The paper focuses on cryptography to secure the info of an organization in motion within its network, at rest, and in usage to unauthorized parties.

1.1 Cryptography

In information, communications and networking, cryptography is important once human action over a transmission medium, particularly the un-trusted medium, significantly the net [5]. The encrypted knowledge is named as ciphertext. At the receiver part, solely those that have a secret key can decipher the message into plain text to obtaining the initial knowledge [6]. Generally encrypted messages are often lessened by cryptography, which is thought as code breaking. Cryptography is often classified into 2 varieties symmetric-key systems and Asymmetric-key systems [3]. In symmetric-key secret writing systems sender

and receiver of the message, create use of the identical key, this distinctive secret is used for secret writing also as coding of the message [3]. In all cases as illustrated in Figure 1, the original unencrypted data is referred to as plaintext (X). It is encrypted into ciphertext (Y), which will in turn (usually) be decrypted into usable plaintext with the same key (K) used for encryption.

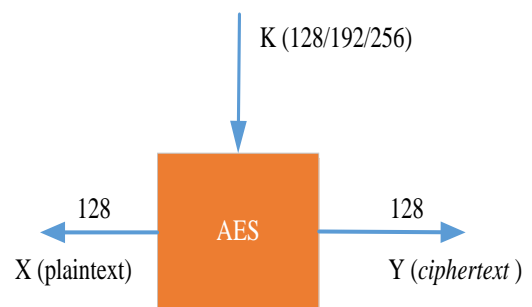


Figure 1 Cryptography Process

Cryptography does not only protect data from theft or alteration, but can also be used for user authentication in most application [5].

2. DESCRIPTION OF AES ALGORITHM

AES is a block cipher which is also known as Rijndael algorithm and it was developed by Joan Daemen and Vincent Rijmen. The algorithm can efficiently be executed on a variety of computer processors and hardware's [7]. AES relies on a style principle referred to as a substitution-permutation network, a mix of each substitution and permutation, and is quick in each software system and hardware. In contrast to its forerunner DES, AES doesn't use a Feistel network [7]. The robust AES development process and its complex internal structures ensures very secure algorithm and has no known weaknesses. In accordance with the AES requirements, Rijndael's key length can be 128, 192 or 256-bits. Rijndael algorithm is made up of variable block size that can also be 128, 192, or 256-bits. This implies that, a Rijndael algorithm with key sizes of 128, 192 and 256-bits provides approximately [8]. AES is one of the most up-to-date out of the four current algorithms approved for federal United States within the US [9]. AES operates on a four \times four (4 \times 4) column-major order matrix of bytes, termed the state, though some versions of Rijndael have a bigger block size and have further columns within the state.

2.1 AES internal structure

Figure 2 shows the overall structure of the AES algorithm, AES composition and building blocks was designed based on standard known as a substitution-transformation arrangement with fixed block size of 128 bits, and a key size of 128, 192, or 256 bits and has a high-speed in both software and hardware. Unlike its predecessor DES, AES is not based on Feistel network [5]. The principle of AES design is known as a substitution-permutation network, which is the combination of both substitution and permutation. AES operates on a 4 \times 4 column-major order matrix of bytes, known as the state. However, some versions of Rijndael have a larger block size and have additional columns in the state. Generally, AES calculations are done in a special finite field called Galois Fields, which allows mathematical operations to scramble data easily and effectively [10] [11]. There are numerous rounds within the AES encryption development. Each operational round consists of more than a few processing steps, each one containing four similar but different stages, including the one that depends on the encryption key itself. The various stages are;

- ✓ ByteSub
- ✓ ShiftRows
- ✓ MixColumn
- ✓ addRoundkey

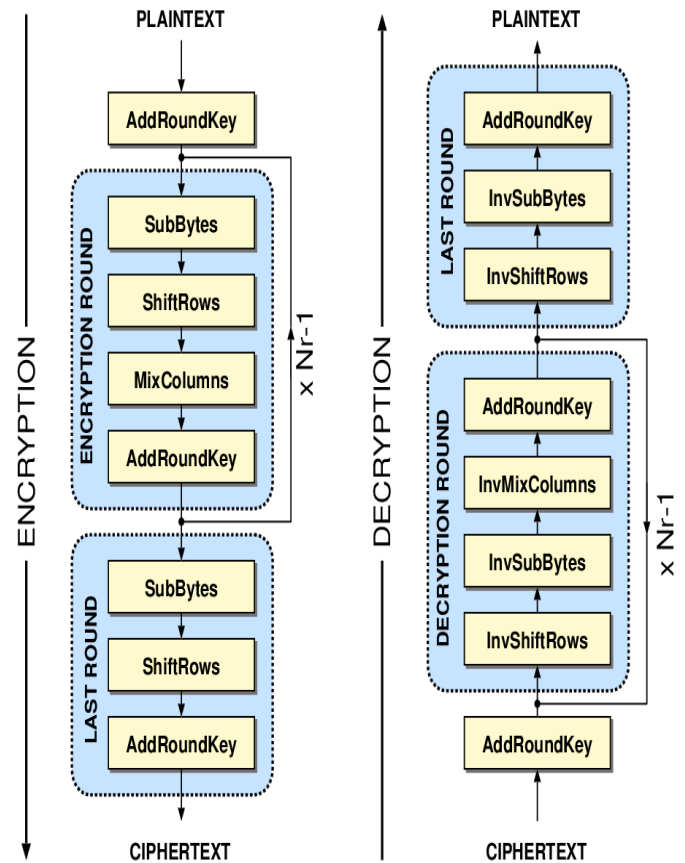


Figure 2 Overall Structure of AES Algorithm (Source: <http://crypto.stackexchange.com/questions/8043/aes-addroundkey>)

2.2 Encryption

The preliminary key is added to the input value at the beginning stage in the encryption mode, which is called a preliminary round. Several repetitions follows immediately after the initial round with a slightly modification of the final. The following operation are perform in every one round respectively.

2.2.1 SubBytes Conversion

The SubBytes convention stage is a non-linear byte replacement, where each state byte is operated independently. This is achieved through an S-box. S-box is a pre-calculated replacement table which holds 256 numbers (from 0 - 255) and their matching resulting value. The SubBytes step has each byte in $a_{i,j}$ in the state matrix swapped with a SubByte $b_{i,j}$ by means of an 8 bits substitution box, known as Rijndael S-box. However, the SybBytes transform algorithm is based on Galois Field Inverse operation $GF(2^8)$ known to have excellent non-linearity properties. The use of Galois Field is to prevent attacks based on simple algebraic properties. The S-box is created by merging the inverse function with an invertible affine transformation. The S-box is selected in order to prevent any fixed operational networks, thus, $a_{i,j} \neq b_{i,j}$. The S-box is created by determining the multiplicative inverse for a given number in $GF(2^8) = GF(2)[x]/(x^8 + x^4 + x^3 + x + 1)$, Rijndael's finite field. Zero, that has no inverse, is mapped to zero. The multiplicative opposite is then changed using the following affine transformation [12]

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Where $[x_7, \dots, x_0]$ is the multiplicative inverse as a vector. This transformation is that the total of multiple rotations of the byte as a vector, wherever addition is that the XOR operation. The matrix development can be deduced using the following algorithm:

Let input number be (s) (thus an unsigned 8 bit variable).

Let the values result be 0.

For 5 times:

XOR result with s.

Revolve s one bit to the left.

Once the matrix exponentiation is done, XOR the value by the decimal number 99 (the hexadecimal number 0x63, the binary number 0b01100011, the bit string 11000110 demonstrating the number in LSB first notation) [12]. This steps will produce an S-box table as shown in Table 1.

Table 1 S-BOX TABLE

		Y															
		0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
X	0	52	9	6a	d5	30	36	a5	38	bf	40	a3	9e	81	f3	d7	fb
	1	7c	e3	39	82	9b	2f	ff	87	34	8e	43	44	c4	de	e9	cb
	2	54	7b	94	32	a6	c2	23	3d	ee	4c	95	0b	42	fa	c3	4e
	3	8	2e	a1	66	28	d9	24	b2	76	5b	a2	49	6d	8b	d1	25
	4	72	f8	f6	64	86	68	98	16	d4	a4	5c	cc	5d	65	b6	92
	5	6c	70	48	50	fd	ed	b9	da	5e	15	46	57	a7	8d	9d	84
	6	90	d8	ab	0	8c	bc	db	0a	f7	e4	58	5	b8	b3	45	6
	7	d0	2c	1e	8f	ca	3f	0f	2	c1	af	bd	3	1	13	8a	6b
	8	3a	91	11	41	4f	67	dc	ea	97	f2	cf	0e	f0	b4	e6	73
	9	96	ac	74	22	e7	ad	35	85	e2	f9	37	e8	1c	75	df	6e
	a	47	f1	1a	71	1d	29	c5	89	6f	b7	62	0e	aa	18	be	1a
	b	fc	56	3e	4b	c6	d2	79	20	9a	db	c0	fe	78	cd	5a	f4
	c	1f	dd	a8	33	88	7	c7	31	b1	12	10	59	27	80	ec	5f
	d	60	51	7f	a9	19	b5	4a	0d	2d	e5	7a	9f	93	c9	9c	ef
	e	a0	e0	3b	4d	ae	2a	f5	b0	c8	eb	bb	3c	83	53	99	61
	f	17	2b	4	7e	ba	77	d6	26	e1	69	14	63	55	21	0c	7d

2.2.2 ShiftRows

In ShiftRows transformation stage, each row of the state is intermittently left shifted above diverse offsets. Row 0 remain in position, while row 1, 2 and 3 are shifted one byte, two bytes and three bytes to the left respectively [13].

2.2.3 MixColumn

From the MixColumn operations, there is a transposition of linear transformation made to join the 4-byte in each column as shown in fig 2.4. The task of this step is to take 4-byte as input and outputs 4-byte, where every input bytes have an effect on all the output 4-byte. Each column is transformed using fixed matrix operations; this is composed of multiplication and addition of the entries as illustrated in fig 2.5. Addition is simply XOR. Multiplication is modulo irreducible polynomial [14] [15]. In the MixColumn process, each column is treated as a polynomial over GF (28) and is then multiplied modulo with a fixed matrix polynomial $c(x)$ multiplies every column, thus,

$$C(x) = 3x^3 + x^2 + x + 2$$

$$\begin{bmatrix} s'_{0,c} \\ s'_{1,c} \\ s'_{2,c} \\ s'_{3,c} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} s_{0,c} \\ s_{1,c} \\ s_{2,c} \\ s_{3,c} \end{bmatrix}$$

$$(0 \leq c < N_b)$$

2.2.4 addRoundkey

For every round in the AddRoundKey step, a subkey is generated from the main key by means of Rijndael's key schedule. The subkey is combined with the state, and that notwithstanding each subkey is the same size as the state. The subkey is inserted by combining every byte in the state with its related byte in the subkey by means of bitwise XOR [16].

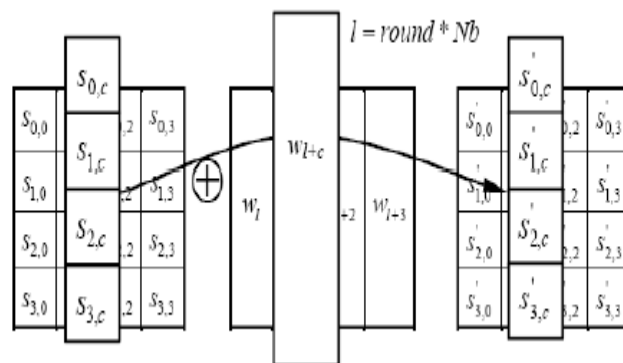


Figure 3 AddRoundKey XORs each column of the State with a word from the key schedule.

2.3 RELATED WORKS

A FPGA-based implementation of the Advanced Encryption Standard (AES) algorithm was proposed by [16]. The design employs an iterative looping technique with block and a 128 bits key size S-box table implementation. The research concluded that low complexity architecture and easily achieves low latency as well as high throughput 1054Mbit/sec for encryption and 615 Mbit/sec for decryption was achieved.

An implementation of high speed AES algorithm with Key Length of 256 Bits based on FPGA is presented was proposed by [17] to advance the security of data in motion. [18] proposed FPGA-based implementation of the Advanced Encryption Standard (AES) algorithm in UART Module. The design employs an iterative looping technique with block and a 128 bits key size S-box table implementation. A throughput 1054Mbit/sec for encryption and 615 Mbit/sec for decryption was achieved in their research. A proposed FPGA-based implementation of the Advanced Encryption Standard (AES) algorithm was also presented by [15]. The design employs an iterative looping technique with block and a 128 bits key size S-box table implementation. Their design was executed using APEX20KC FPGA on Altera which is based on great performance design. In their paper, a low latency and the throughput attained a value of 1054Mbit/sec for encryption and 615 Mbit/sec for decryption [15]. An implementation of AES encryption and decryption standard AES-128 was proposed by [19]. All the transformations of each secret writing Associate in nursing decoding are simulated victimization an unvarying style approach so as to reduce the hardware consumption. Their paper proposed that their methodology will create it a really low-complex design, particularly in saving the hardware resource in implementing the AES InverseSub Bytes module and Inverse combine columns module. Because the S - box is enforced by look-up-table during this style, the chip space and power will still be optimized. The new combine Column transformation improves the performance of the inverse cipher and additionally reduces the quality of the system that supports the inverse cipher. As a result this transformation has comparatively low relevant diffusion power .This allows for scaling of the design towards vulnerable moveable and cost-sensitive communications devices in client and military applications [19]. An implementation of the 128 bit AES on a Field Programmable Gate Array (FPGA) was proposed for significant level of security similarly as quicker time interval in order that it will be used for secured communication of ATM, optical disc content similarly as for secured storage of confidential company documents, government documents [3].

The above discussions show that an implementation of AES algorithm with 128 key length can be enhanced with a 256 bits key length to provide a well secured data [9]. Rijndael's reduced variety of rounds for smaller keys offers it a speed advantage, however it additionally reduces the security issue for those key sizes [7], hence this paper focuses on the implementation of the AES algorithm with a 256 bits key length to prevent data loss in an organization.

3. TOOLS AND METHODS

Microsoft Visual Studio 2012 (C#) was employed to develop the face, where system users (client) will diagrammatically communication to the server via a web browser. The rear finish (database) was developed with Microsoft Structured command language (MSSQL) server 2008. A Wireshark hacking tool was used for testing the encrypted data.

3.1 Design concept

Figure 4, shows the pictorial view of the proposed layout for the implementation of the AES algorithm with 256bits key length for organizational data protection. Records entered by the organization system users from the client machine from remote stations are encrypted using the ASE 256bit algorithm

on the remote station and sent or transmitted to the data centre (server) true the clouds (internet) or LAN and MAN.

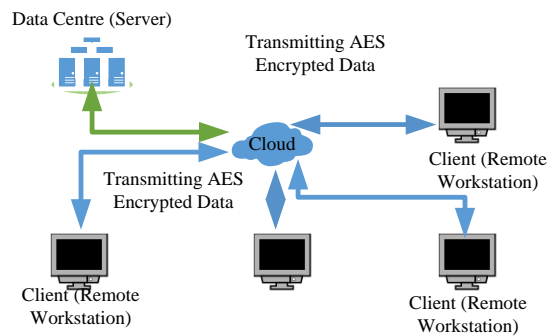


Figure 4 Conceptual Design of Proposed System

For implementation, a web application with login details of client as shown in figure 5 was developed to collect raw data in plaintext to be encrypted by the AES for testing the outcome of the implementation of the AES 256bits algorithm.

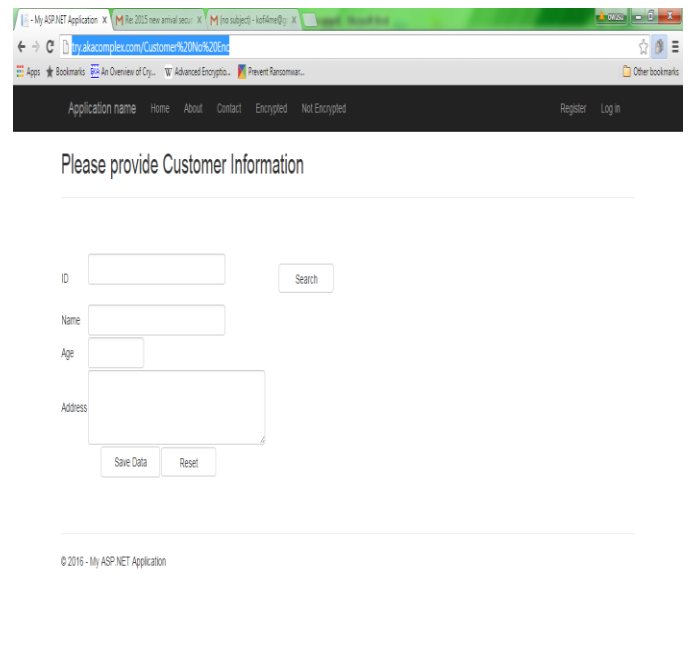


Figure 5 Plaintext data collection interface

4. TEST RESULTS AND DISCUSSIONS

To see the effectiveness of our AES 256bits implementation, five (5) different data was collect from the front end. The data transmission from the client machine to server was done in two scenario A and B. In scenario A the data was transmitted raw (no encryption) over vulnerable medium to the server. The Wireshark hacking tool was used to intercept the data packets in transmission and the raw format of the packets are displayed by the network packet analyzer in a plain text as illustrated in figure 6. This shows transmission on unencrypted data over a network is exposed to fallen into the hands of unauthorized persons.

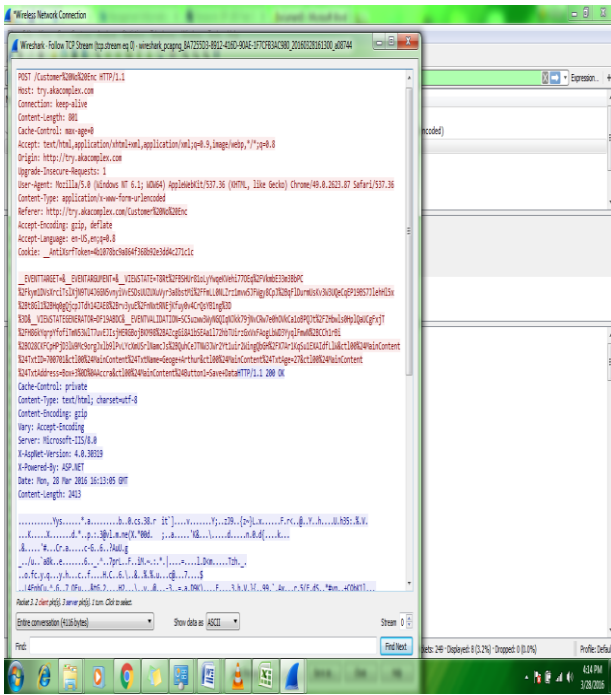


Figure 6 Unencrypted data of the client with ID: 700701

Same client with ID 700701 data was transmitting under scenario B, which fully encrypted with AES 256 bits block cipher before transmission. Again the Wireshark hacking tool was used to intercept the packets in transit but the network packet analyzer displayed scrambled letters which has no meaning to the hacker as shown in figure 7.

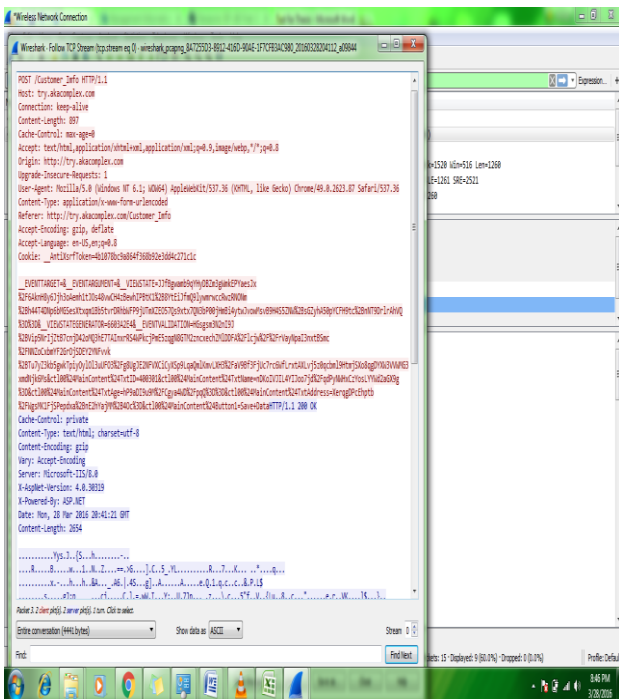


Figure 7 Encrypted data of the client with ID: 700701

The two scenario A and B sent data over a transmission medium to a dedicated server. In scenario A, we were able to transmit data without encrypting it, hence the data stored in the server was in plaintext. Whereas in scenario B, an AES

256bits key length was used to encrypt the data before transmitting to the server. The aim of this research was achieved in in scenario B, thus encrypting data with AES 256bits key length before it is transmitted to the server. An organization that do not deploy encryption in its daily business transactions or data transfer stand a higher risk of losing very critical and sensitive information hackers as seen in scenario A. The results from the unencrypted data were obvious, as all the input data were transmitted in it plaintext format to the server. This possess a great danger to the organization, especially about the data of its customers and partners, e-commerce and other business transactions.

Regarding scenario B, the input data was scrambled by the AES 256bits key length algorithm before transmission. When malicious people intercept the data, because it has been scrambled by the AES 256-bit algorithm, they cannot read the ciphertext unless they crack it first. For a hacker to be able to crack the key of the encrypted data, it will require of him enormous tasks that span across several period of time, thus, 256 bits AES application: $2^{256} = 1.1 \times 10^{77}$ possible keys. It is assumed an attacker with the capability to build or purchase a system that tries keys at the rate of one billion keys per second requires many centuries to crack AES. With this hypothesis, the attacker would require about 1×10^{21} years to try all likely keys for the weakest edition of AES-256.

Since the application uses AES encryption and decryption key size 256 bits and a cipher block size 256 bits, the ciphertext is so strong that it becomes virtually impossible for any malicious person to decrypt it. Since the ciphertext was encrypted with AES 256-bit and key size of 256 bits, a bad person will requires 2256 possible keys to be able to use brute force to decrypt a character of the key.

4.1 Results for unencrypted data in the server

Table 2 shows the stationary data saved in the server which clearly indicates the system did not implement any form of encryption which means all the data saved are in plaintext under section (A) category. It is very risky and extremely dangerous to transmit sensitive information without proper measure to protect or shield it. This implies that any hacker getting access to the save can easily get hold of the organization data in its raw (plaintext) format.

Table 2 Unencrypted data saved in the server

SQL \SQL019 \Databases \DB_964804_ajacomplex \Tables \DB_964804_ajacomplex \tbl0 \Contacts

Query: SELECT * FROM [DB_964804_ajacomplex].[tbl0].[Contacts]

Record(s) found: 5

ID	UID	FullName	Age	Address
70	700701	George Arthur	27	Box 3 Accra
71	700702	Ahlye Odebe	29	Box 6 Sunyani
81	700703	Dominic Mesh	30	Box 82 Kumasi
81	700704	Michael Asante	37	Box 73 Adansi
82	700705	Onusa Boabeng	32	Box 83 Koforidua

4.2 Results for encrypted data in the server

Table 3 shows the saved data of an organization that transferred data in an encrypted. The input data, the intercepted data and the data saved in the server are different. The input data has been scuttled by the AES 256 bits block cipher algorithm hence, the saved data in the server has been completely protected by the application.

Table 3 Encrypted data saved in the server

The screenshot shows a SQL query result in a database interface. The query is: `SELECT * FROM [DB_964844_akacomplex]([dbo])Contacts`. The result set contains 5 rows of data, all of which are heavily encrypted. The columns visible are ID, VNO, FullName, Age, and Address. The data in these columns is represented by long strings of alphanumeric characters, indicating that the original data has been encrypted.

ID	VNO	FullName	Age	Address
87	40301	r2KzL0L4Y12oo7d qPBYW4H+Cz1rel.YVW2a3G9g=	hP8z05hBM Cqy#41D oqQ==	lveqDRe5tpd WgMKJFSPepta+Z2h1gM+4Qz=
89	40302	SgrHyZ0GEFevr8e38EMV5Q+8R0J#R0qz=	0lzc1VE5L0B ReACT ggr==	U4C Ov 8jAV7 UG0 RR on7mB W CZLLS8Ev=
91	40303	eeCzG2W8R0S W+84C74q CZLL+H U8B8 8=	69eeUG0D 5e4eC hIn v==	vU4Hve8a U86Cm NPLuB7OO C D gD2e+5=
93	40304	aFC0X44+080G u 7 553K 5P8 + W PMy Z8H h=	3y#8Z2Z8E8 Q rot UL5v ==	HFGQ h g p 8P 8 L H B W v C C q g 8 C v=
95	40305	7AZUk2K5u G W Z U N V G Z U Z 8 8 D 8 U C 3 P h Q z=	cTg5 8e g N D Hery v ==	q K W 7 8 F V Z C U D D e j c 8 8 F E W 8 6 W U U E=

5. CONCLUSION AND RECOMMENDATION

The conclusions of this research could be summarized as follows:

- ✓ The tests conducted indicate a high level security for data in transit and stationary. The data protection using AES is to provide optimized data security to classified and non-classified data. The test conducted for all the five client (5) using the developed web application indicated a successful data protection. When encryption algorithm was used to encrypt the data, it was realized that data in transit, data saved in the server are highly protected and major data losses.
- ✓ The results also results of the data seen in the server shows that even if a hacker hacks or intercepts the data through a hacking tool or social engineering, the data will be meaningless the him/her. This research will solve problem of organizations losing their sensitive data to unauthorized persons, a 256 bits key length offers more security to data both at rest and in transit [3].
- ✓ Organizations with virtual offices at remote location require this form of application to enable their remote workstations to communicate securely with the server. Since the ciphertext was encrypted with AES 256-bit and key size of 256 bits offers a better and a more secured that as compared with [19] [15] [16] [16] [3] that employed 128 key length, anyone who want to crack an AES with 256 key length will requires 2^{256} possible keys to be able to use brute force to decrypt a character of the key. An average throughput 1054Mbit/sec for encryption and 615 Mbit/sec for decryption was realized from the test analysis.

Recommendation

Due to the high level security capabilities provided by AES with 256 bits key length, we recommended to organization such as school, banks, microfinance and churches, that seeks to do a secure business transaction both online or on corporate network infrastructure to protect the data of its clients, staff, partners and supplier with AES with 256 bits key length algorithm.

6. FEATURE WORKS

Even though the AES algorithm offers data security and better encryption as shown by this paper, we wish that researchers who wish to extended and improve upon this work should look at combining two encryption algorithm such as AES and DES to give much and better organization data security and protection from unauthorized persons and hackers, since these two methods have unique features.

REFERENCES

- [1] Ernst & Young, "Data loss prevention: Keeping your sensitive data out of the public domain," Ernst & Young Global Limited, UK, 2011.
- [2] A. Ibrahim, "FPGA-based Hardware Implementation of Compact AES Encryption Hardware Core," *WSEAS transactions on circuits and systems*. ISSN: 2224-266X, vol. 14, 2015.
- [3] B. ., Madhuri and N. M. Suresh, "Implementation of Advanced Encryption Standard Algorithm for Communication Security Using FPGA," *International Research Journal of Engineering and Technology (IRJET)*, vol. 03, no. 07, pp. 1176-1179, 2016.
- [4] F. Twum, k. Nti and M. Asante, "Improving Security Levels In Automatic Teller Machines (ATM) Using Multifactor Authentication," *International Journal of Science and Engineering Applications*, vol. 5, no. 3, pp. 126-134, 2016.
- [5] G. C. Kessler, "An Overview of Cryptography," 2008. [Online]. Available: <http://www.garykessler.net/library/crypto.html>. [Accessed 27 January 2015].
- [6] V. Beal, "cryptography," 2014. [Online]. Available: <http://www.webopedia.com/TERM/C/cryptography.html>. [Accessed 14 June 2015].
- [7] B. Schneier, J. Kelsey, D. Whiting, D. Wagner, C. Hall, N. Ferguson, T. Kohno and M. Stay, "The Twofish Team's Final Comments on AES Selection," 15 May 2000. [Online]. Available: <https://www.schneier.com/academic/paperfiles/paper-twofish-final.pdf>. [Accessed 2 February 2016].
- [8] W. M. Tatum, "The Advanced Encryption System (AES) Development Effort: Overview and Update," SANS Institute, 2001.
- [9] M. Pitchaiah, D. Philemon and Praveen, "Implementation of Advanced Encryption Standard Algorithm," *International Journal of Scientific & Engineering Research*, vol. 3, no. 3, pp. 1-6, 2012.

- [10] J. Nechvatal, E. Barker, L. Bassham, W. Burr, M. Dworkin, J. Foti and E. Roback, "Report on the Development of the Advanced Encryption Standard (AES)," Computer Security Division Information Technology Laboratory National Institute of Standards and Technology Technology Administration U.S. Department of Commerce, U.S.A, 2000.
- [11] T. Alaa, A. Zaidan and B. Zaidan, "New Framework for High Secure Data Hidden in the MPEG Using AES Encryption Algorithm," *International Journal of Computer and Electrical Engineering*, vol. 1, no. 5, pp. 1793-8163, 2009.
- [12] Wikipedia, "Rijndael S-box," 11 November 2013. [Online]. Available: https://en.wikipedia.org/wiki/Rijndael_S-box#cite_note-1. [Accessed 12 January 15].
- [13] G. Mohan and K. Rambabu, "An Efficient FPGA Implementation of the Advanced Encryption Standard Algorithm," *International Journal for Scientific Research & Development (IJSRD)*, vol. 2, no. 07, pp. 413-417, 2014.
- [14] C. Paar and J. Pelzl, "Advanced Encryption Standard," 2009. [Online]. Available: https://en.wikipedia.org/wiki/Advanced_Encryption_Standard. [Accessed 25 May 2016].
- [15] T. Hoang and V. L. Nguyen, "An efficient FPGA implementation of the Advanced Encryption Standard algorithm," *In Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), IEEE RIVF International Conference*, pp. 1-4, 2012.
- [16] S. ., R. Kumar and V. Viswanadha, "An efficient FPGA implementation of the AES Algorithm With Reduced Latency," *International Journal for Scientific Research & Development (IJSRD)*, vol. 1, no. 10, pp. 2074-2077, 2013.
- [17] K. Gayathri and W. Yasmeen, "Data Encryption and Decryption using AES with Key Length of 256 Bits," *International Journal of Scientific Engineering and Technology Research*, vol. 03, no. 20, pp. 4143-4146, 2014.
- [18] C. Sadashiva and S. Sunkari, "Data Encryption and Transition by AES Algorithm with UART," *International Journal of Scientific Engineering and Technology Research*, vol. 03, no. 35, pp. 6935-6938, 2014.
- [19] P. Aatheswaran and R. Babu, "FPGA can Be Implemented by Using Advanced Encryption Standard Algorithm," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 1, pp. 675-679, 2013.
- [20] P. Anitha and Palanisamy, "Data Protection Algorithm Using AES," *International Journal of Current Research*, vol. 33, no. 6, pp. 291-294, 2011.
- [21] N. Ahmad, R. Hasan and W. Jubadi, "Design of AES S-Box using combinational logic optimization," *IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, pp. 696-699, 2010.

Evaluation of Server Performance in Cloud Computing On Virtual Network Infrastructure: Using Riverbed Simulator

Charles Amo-Asuah
T. I Ahmadiyya SHS
Kumasi, Ghana

Isaac Kofi Nti
Electrical/Electronic
Engineering Department
Sunyani Technical University
Sunyani, Ghana

Samuel Akyeramfo-Sam
Department of Computer
Science
Sunyani Technical University
Sunyani, Ghana

Abstract: Computing in the cloud changes the way **Information Technology** (IT) is managed and disbursed, given an improved cost-efficacies, faster novelty, quicker time-to-market, and the capacity of scaling applications on request (Arora & Gupta, 2012). Most organizations are now opting for web based services, since the use of virtual networking infrastructure in cloud storage does not only reduce costs, eliminates hardware failures and security risks such as theft of physical components in the real world, but also provides a graphical user interface for the topology design allowing for realistic simulation of networks and high performance displaying modules as an added advantage for modern enterprises. This research focuses on examining the performance and relations of distributed systems in servers used for cloud storage on virtual platforms, by analyzing the behavior and efficiency of server performances in three different scenarios in a cloud network using Riverbed Edu 17.5 edition as a virtual network platform simulation tool. Networks with three different server scenarios simulated against time in a network were modelled. The focus was on the server performance, at different loads and processing speeds. The simulation was configured for about 100 workstations in a manner that all the applications and users can access the parameters; database, file transfer protocol, hypertext transfer protocol, and email response times on cloud utilization throughputs in two directions. The results for server performance in efficiency and viability is analyzed and evaluated with conclusions drawn as a guiding principle for organizations and companies considering the use or using servers for cloud storage.

Keywords: Cloud-Computing; Networking; Servers; file-transfer-protocol; cloud-utilization; throughputs; email-response-time; Virtualization

1. INTRODUCTION

Computers have grown with rapid changes in technology, and they have become increasingly faster than ever to the extent that users of microcomputers are said to be under utilizing the potentials of their computer systems both at home and in the offices or work place. This paper centers on virtual network presented as part of cloud service as a model of virtual resource interconnect, where resources are dynamically deployed and migrated in a cloud. Visualization stands for pulling out hardware from existing software. The process of extraction can be done in many forms, by the use of operating systems such as Hyper-V, creating visualizations via terminal systems or by means of App-V and with establishments such as Hewlett Packard and Cisco which runs both network and visualization storage. Organization globally have incorporated the visualization operating systems in the workstations to augment the delivery and the performance of their day to day activities (Desai, 2012). Virtualization can be defined as the act of designing the real virtual multiplicity of an item such as device storage, operating system (OS), resource network and hardware platform with the physical outlook of a computer representing a whole machine incorporating and encompassing both subjective and objective point of view in a non-questionable manner (Turban, et al., 2008). Subjectively, the virtual electronic gadget is whole or complete whereas from an objective perspective, they are just a collection of files which operates on a physical machine. Globally, the virtualization technology is widely accepted and has become part of daily life activities, owing to its attractiveness. The virtualization technology operational techniques permits easy distribution of information between distinct virtual machines (VM), by this means enhancing the maximum use of hardware's. The flexibility with the use of virtual machines makes its management easier and has also spawned new models of efficiently operating the system (Koponen, et al., 2014). In a more sensitivity view, virtualization is vital in operating cloud

computing minus which cloud computing would appear anomalous. Virtualization has made IT cheaper and a global commodity. The solid foundations of cloud computing is built on the prowess of virtualization which has cost effective use of the model with emphasis on scalability and the rate at which the service is demanded (Turban, et al., 2008). The cloud system provides a common platform for easy access, managing and operating resources. Also cloud computing constitute the tendency to use scalable and shared IT platform within the parameters of an internet (Kaufman, 2009). Virtualization technology is the heart of cloud computing. Unlike other archetypes, cloud computing is yet to gain a common grounds regarding its definition due to the fact that its parameters are frequently varying and increasing. However, the US National Institute of Standards and Technology (USNIST) have provided a concise definition as; "Cloud computing is an IT architecture that provide more convenient services to user on their computers which requires minimal supervision from provider to operate (Mell & Tim, 2009). There is the existence of resource sharing among virtual networks. This means that if each product draws on separate protocols, there is the need for discrimination to maintain integrity within the cloud system. Moreover, in order to deliver quality service, adapting bandwidth is important due to the fact that different approaches are required for different bandwidth. There are some IT developers such as Apple, Google and among other who have develop their own cloud systems to enhance user preference. With the problems associated in the use and adoption of traditional methods of network storage and hosting, this paper seeks to assess the technologies, advantages, cost-efficacy and deployment competence of virtualization infrastructure employed in cloud computing systems in today's computing over traditional platforms.

2. TYPES OF VIRTUALIZATION

There are numerous types of virtualization in computing but this paper seeks mention and explain two of them.

2.1 Server Virtualization

Virtualization is the main and primary accelerator of cloud-computing and cloud-storage (Arora & Gupta, 2012). Virtualization permits the run of multiple applications on a virtual machine within the same physical server as a substitute to running only one application on a physical server. It enable delivering of greater accessibility, scalability, and optimization of resources such as servers and storage (Arora & Gupta, 2012). Carefully analysing Arora & Gupta definition of server virtualization, one can say that the server is said to have so many resources that are available for users to use. Not ever said resources is easily and user-friendly for the virtual IT environs when serving multiple isolated virtual environment with single server.

2.1.1 Types of Server Virtualization

(Baburajan, 2011) states that, in order to virtualize resources or materials, there should be virtual machine (VM) models, OS stage virtualization as well as Para-virtual machine models. The research analyzed and chose amongst alternatives. There are different types of server virtualizations on the IT markets. Management has to decide on the one which best suits the institution's processes in order to achieve the much needed efficiency gains from server virtualization most importantly on the cost saving advantage that comes along with virtualizing systems.

2.1.2 VirtualMachine Based Virtualization

The virtual machine virtualization server runs on virtual machine monitor or can also use the hypervisor in transmitting command to and receiving information from the CPU and this is done without any extra modifications. Subsequently, VM (Virtual Machine)-based Server Virtualization give protection to the hardware layer, partition can be done within the server while running on separate operating systems without any guest application systems and knowing that they are all running on a mother OS (Saunders, et al., 2009)

2.2 OS Level Virtualization

Operating System level Virtualization is attained through the mother OS running on an isolated system and through the management and controlling of guest OS functionalities. This means that virtual level operating systems draws from the creation of segregated containers in unified physical servers while using each segregated entities to work separately of the partitions (Baburajan, 2011).

Types of Desktop Virtualization

For this research purpose four types of VDI operation are stated namely;

- ✓ Hosted (mostly used in the form of a server)
- ✓ Centralized
- ✓ Remote Synchronization
- ✓ Client-Hosted

In order to sustain the Hosted as well as the centralized modes, there is the need for the continuous supply of uninterrupted network when connected to server with the VDI running. This technological technique is in line with the concept of thin clients which also depends on uninterrupted internet service whereas the remote synchronization allows individuals to copy VDI system and operate it on another machine without connection. With this model individuals access images from central server which is copied and later used on personal computers when making a journey. This form of uninterrupted access to and working on work pages at anywhere and at all times have their own merits and demerits.

2.3 Benefits and Drawbacks of Virtualization

- ✓ There are numerous benefits to be derived from server and desktop virtualization:
- ✓ It reduces solid containers (servers), hence cutting down cost and maintenance overheads.
- ✓ More space are created with an environment with less physical hardware's.
- ✓ Spontaneous performance of multiple task on one platform.
- ✓ Risk of losing data is less.
- ✓ Allowing employees to work using a common platform.

Drawbacks

- ✓ In desktop Virtualization, when the server goes down all users/client can't work.
- ✓ There is lack of user privacy, thus since all the processing is done by probably less or single servers remotely, everything or any task that a connected terminal does can easily be monitored.
- ✓ Sometimes users are restricted as to the number of tasks they can perform on the terminals or virtualized systems.

2.4 Cloud Technology and Cloud Computing

The term cloud computing is a new paradigm in the field of IT that has attracted attention from researchers, organizations such as the banking industry, educational institutions, SMEs among others (Luis, et al., 2008). This model talked about is expanding at a faster rate and attracting various vendors and clients. The faster expansion of cloud computing is being geared by the emanating technologies which gives a clear and rational price use of computing structures and aggregate storage capacity. It lessens the need for huge investment in information Technology infrastructure. This computing model implicates procuring of computer properties with the ability of increasing resource capability of being scaled, on-demand supplying of less no leading IT cost of investing in infrastructure. The advantages in cloud computing is offered through three delivery models called infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and software-as-a-service.



Figure 1 Cloud Server on virtual desktop

2.5 Cloud Computing Services

Cloud computing services can be grouped into three categories:

2.5.1 Software as a service (Saas)

In delivering cloud computing as a service, users patronize the right to access the service for being hosted on the cloud system. Currently, Saas is being offered to individuals and organizations by IT giants such as Google, Microsoft, Sales force and among others. The system gives responsibility to users to patch and maintain the OS and the applications software.

2.5.2 Platform as a Service (Paas)

Paas offers cloud users the opportunity to access platforms as well as problem resolving stacks such as programming language, database, web servers and operating systems. Cloud users have the liberty to develop their own applications which operates on the cloud server. Among the popular Paas used are Google App engine, Force.com and others.

2.5.3 Infrastructure as a Service (Iaas)

Iaas offers fundamental storage and computing, which is done to standards as against the provision of network. Organizations and individual users are allowed to maintain and control their own applications, issues regarding storage, operating systems and network connections. Examples of Iaas are as follows; Amazon, GoGrid, 3 Tera etc.

3. MATERIALS AND METHODOLOGY

Riverbed Edu 17.5 was used as a simulation tool

Riverbed Edu.version17.5 was used as a simulation tool, it offers virtualization for development, analyzing and evaluating the capacity and performance of virtual systems which includes application servers and network topologies.



Figure 2 Riverbed Edu 17.5 as a Simulation Tool

4. Results and Discussions

Three different scenarios were implemented with the simulation tool and tested for an hour on 30mb/s, 2gb/s and 10gb/s loads for 100mb/s bandwidth in each case of the experiments.

In the first scenario, the server was used in the network with four applications and specified number of users (10-60) passing through the cloud and server to generate the required traffic across the system. The second and third instance, the servers were two and three respective with the same applications and users to generate the required traffic across the distributed system.

The email, ftps web application and database executions were assessed and the output parameter measured for all three levels, cloud/router measurement and connection measurement respectively. Each diagram acquired was analysed against the execution measurements and a critical scrutiny was done.

4.1 Database Application Results

The database application generated a lot of traffics when measured against the response time

Database Query Response Time:-This is the slipped by time between the end of a request, inquiry or interest on database server and the start of a request.

4.2 Database Query Response Time–One Server

The tables 1 and 2 shows the database query response time when one server was implemented and figure 2 gives the combined graph in Figure 3 showing the graphical representation and performance of the one server scenario.

Table 1 Database query response time with packed size of 50MB (low)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	0.26	0.27	0.28	0.27	0.22	0.23
Double server	0.24	0.21	0.27	0.21	0.23	0.26
Triple server	0.25	0.22	0.25	0.22	0.36	0.36

Table 2 Database query response time with packet size of 250MB (High)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	1.94	2.36	2.29	2.36	2.38	2.71
Double server	4.43	6.79	4.43	3.13	3.98	4.09
Triple server	2.42	2.28	2.61	1.97	22.53	16.65

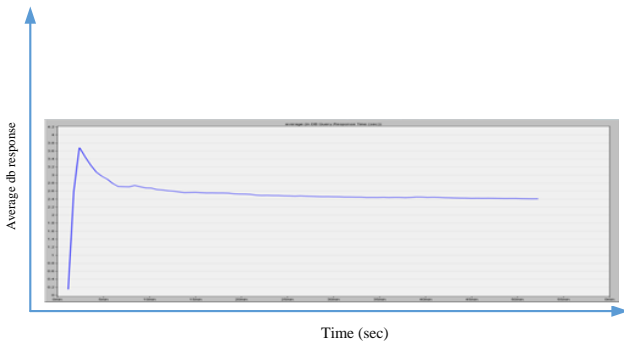


Figure 3 Database Query Response Time–One Server Scenario

4.3 Database Query Response Time–Two Server Scenario

Tables 1 and 2 row 4 represents the performance of the two server scenario, and the combined graph in Figure 4 shows the graphical representation and performance of the two server scenario.

4.4 Database Query Response Time–Three Server Scenario

In the third scenario, another server was introduced. The tables 1 and 2 Row 5 shows the response time when other applications were diverted to other servers on the network by the LAN.

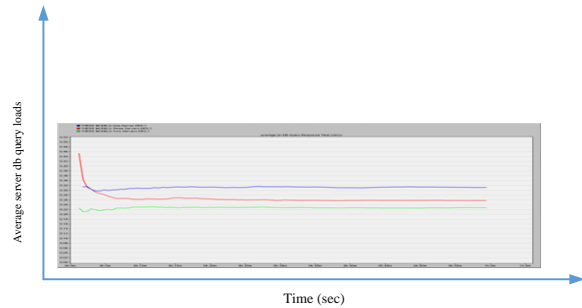


Figure 4 Database query loads

From Figure 3, gives the response time of databases query when other applications were diverted by LAN to the three servers within the cloud network. It can be seen that the traffic across the network had an initial low value of 0.18secs and then increased to 3.37secs for a packet size of high (250mb). It then declined from 3.37secs to 2.24secs and then maintained a constant value throughout the simulation period. Figure 4 shows the combined graph of all three scenarios. The differences in the effect is as shown in the graph.

4.5 Server Database Query Load

Server database load is the time taken for the database server to receive data from users across the network, process it back and forth across the cloud network.

4.6 Server Database Query Load–One Server Scenario

The values in tables 3.3-3.4 showed the server loads across the network when only one server was used. Refer to the combined graph in Fig4.4 for the graphical representation and performance of the one server scenario.

4.7 Server Database Query Load–Two Server Scenario

The result in tables3.3- 3.4 showed the load on the server when two servers are used on the network.The combined graph in Fig4.4 showed the graphical representation and performance of the one server scenario.

4.8 Server Database Query Load–Three Server Scenario

In the third scenario, all the four applications were passed through their destination servers and the load analyzed and evaluated on the cloud network. The tables 3 and 4 showed how the loads performed in the network.

Table 3 Server database load with packet size of 50MB (Low)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	0.72	0.94	0.72	0.94	0.44	0.58
Double server	0.53	0.58	0.53	0.58	0.36	0.50
Triple server	0.69	0.69	0.69	0.69	0.33	0.58

Table 4 Server database load with packet size of 250MB (High)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	1.77	1.81	0.06	0.06	2.08	2.08
Double server	2.64	2.64	0.06	0.06	2.10	1.25
Triple server	2.18	2.05	0.00	0.08	0.69	0.78

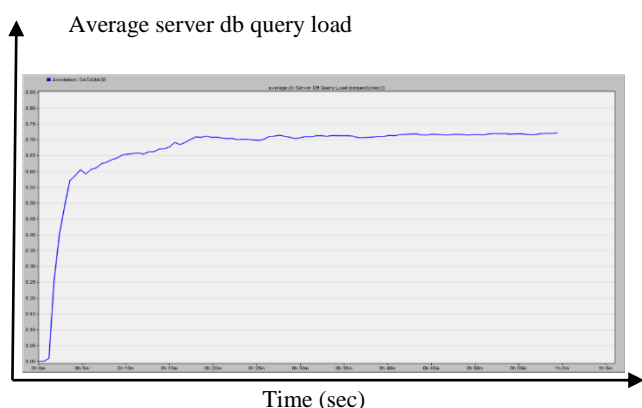


Figure 5 Server Database Query Load–Three Server Scenario

Figure 5 shows that, the server database query load had an initial load of 0.00secs and later increased to a high value of 0.725 secs

and retained that value throughout the simulation period at all packet sizes and data rates and in all three servers. The result of combined graphs for the three scenarios is as in Figure 5.

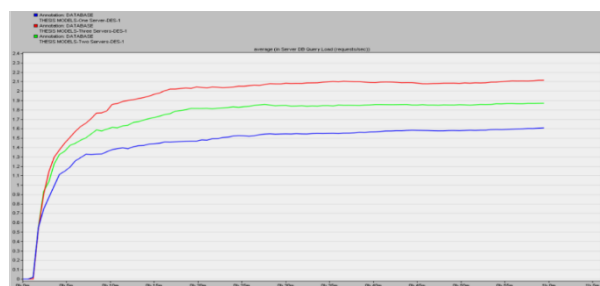


Figure 6 Server Database Query Load

From Figure 6, it can be seen that the load on servers is almost equal in all packet sizes. It could be seen that, the load on the server one is less than on servers two and three. It took more time to process user requests in the second and third scenarios because more time was spent to route applications to their destination servers.

4.9 Results for E-mail Application

Server performance was analyzed and evaluated in this section against the email downloads and uploads response time when the three scenarios were considered. E-mail with packet size of low (50mb), medium (150mb) and high (50mb) were used in link speeds of 10gbps, 2gbps and 30Mbps configured and evaluated against the performance metrics.

4.10 E-mail Download Response Time–One Server Scenario

The tables 5 and 6 shows the e-mail download response time when one server was used on the network and the combined graph in Figure 8 shows the graphical representation and performance of the one server scenario.

Table 5 E-mail downloads response time with packet size of 50MB (Low)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	0.59	0.80	0.59	0.08	0.71	0.73
Double server	0.76	0.99	0.76	0.99	0.66	1.03
Triple server	1.03	0.97	1.03	0.97	0.87	0.92

Table 6 E-mail downloads response time with packet size of 250MB (High)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	1.27	1.23	1.27	1.23	1.10	1.16
Double server	1.42	1.19	1.44	1.23	2.16	2.01
Triple server	1.31	1.24	1.33	1.89	5.08	4.70

4.11 E-mail downloads responses time–Two Servers Scenario

The result of the download response time for two servers was used on the network are shown in tables 5 and 6. The result shows an increase in downloads response time when data packet sizes increased.

4.12 E-mail downloads response time–Three Server Scenario

In the third scenario, the email application was diverted to its preferred destinations in the LAN across the routers and cloud to the servers in the network. Table 6 showed the results when email application was analyzed and evaluated against the performance metrics among the three servers in the network.

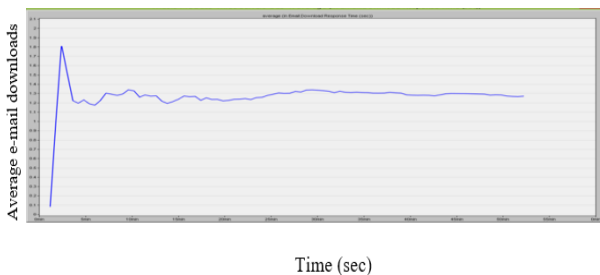


Figure 7 E-mail downloads response time–Three Server Scenario

From Figure 7, the e-mail response time which is the time taken for the web server to process user requests across the cloud network back and forth, was evaluated against the performance metrics within the three servers. From the graph, it could be deduced that the load had a low value of 0.8secs at high (250mb) and medium (150mb) packet sizes to 1.8secs and then declined to 1.2secs and maintained a constant value throughout the simulation period. It was concluded that the e-mail response time was high when the load on the network was high and vice versa. The combined graph in Figure 8 shows the graphical representation and performance of the three scenarios.

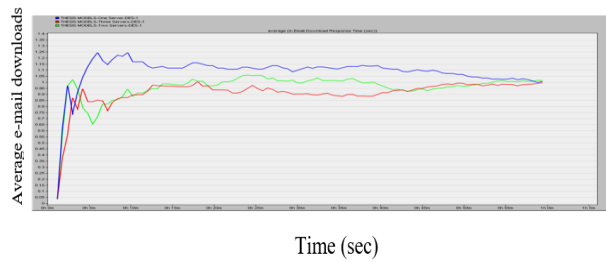


Figure 8 E-mail downloads response times

Figure 8 gives account of all three scenarios for e-mail downloads response times.

4.13 E-mail Upload Response Time

This is the response time taken for data to be uploaded unto a cloud network.

4.14 E-mail Upload Response Time–One Server Scenario

The upload response time was evaluated in this section. The tables in 7 and 8 shows the upload response time when one server was used on the network. The graph in Fig 4.8 showed the performance of the one server scenario.

4.15 E-mail Upload Response Time–Two Servers Scenario

The tables 7 and 8 shows the values when two servers were implemented on the network. A packet latency of 0.05secs was enforced to introduce delay into the system. Refer to the combined graph in Figure 10 for the graphical representation and performance of the two server scenario.

4.16 E-mail Upload Response Time–Three Servers Scenario

In the third scenario, the email application was routed by the routers to the three servers on the network. The results in tables 7 and 8 shows the response time when email was uploaded across the cloud network.

Table 7 E-mail Upload Response Time with packet size of 50MB (Low)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	1.19	1.00	1.19	1.00	0.45	0.79
Double server	0.55	0.74	0.60	0.74	1.04	1.12
Triple server	1.08	0.93	1.08	0.93	1.43	1.52

4.17 E-mail uploads response time for the Scenarios

It is the e-mail response time received by the server when users make a request to the e-mail server back and forth across the network. Combined effects of the three scenarios is as represented in Figure 10.

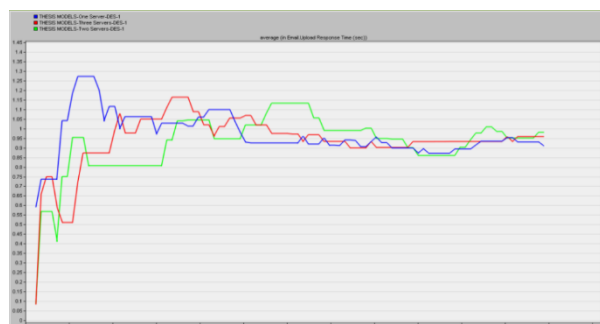


Figure 10 E-mail Upload Response Time

Figure 10 shows the simulation results when all the three scenarios were considered. It can be seen from graph that the upload response time was high when one server was used on the network than when two and three servers were used.

Table 8 E-mail Upload Response Time with packet size of 250MB (High)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	1.16	1.39	1.56	1.39	1.29	1.14
Double server	1.12	1.35	1.11	1.27	3.34	2.61
Triple server	1.48	1.46	1.38	1.42	4.85	4.89

4.18 Cloud Utilization Performance

Cloud utilization was evaluated against the point to point utilization. Network utilization is the proportion of current network traffic to the maximum traffic that the port can work with. It indicates the bandwidth use in the network. While high network use indicated the network was busy, low network utilizations indicated the network was idle. At the point when system usage surpassed the limit under ordinary condition, it would cause low transmission speed, intermittence and request delays. The tables 9 and 10 shows the link utilization for applications on the network and Fig 4.18 showed the link utilization across the three scenarios

Table 9 Cloud utilization with packet size of 50MB (Low)

Scenario s	10gb/s		2gb/s		30mb/s	
	10 min s	50min s	10min s	50min s	10min s	50min s
Single server	0.38	0.30	0.31	0.31	0.32	0.23
Double server	0.32	0.26	0.22	0.24	0.38	0.28
Triple server	0.36	0.29	0.26	0.21	0.54	0.47

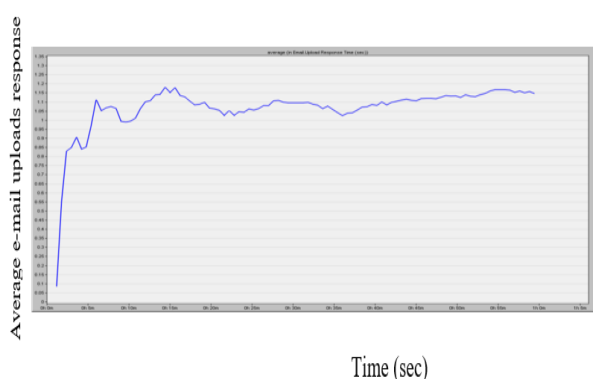


Figure 9 E-mail uploads response time–Three Servers Scenario

From Figure 9, e-mail upload response time had an initial low value of 0.008secs to a high value of 1.155secs due to routing protocols before reducing to an appreciable value of 0.05secs and then maintained a constant value throughout the simulation period.

**Table 10 Cloud utilization with packet size of 250MB
 (High)**

Scenario s	10gb/s		2gb/s		30mb/s	
	10 mins	50mins	10mins	50mins	10mins	50mins
Single server	12.74	12.56	11.12	12.56	1.28	1.56
Double server	19.63	12.49	21.27	13.31	1.45	1.30
Triple server	39.78	38.16	44.76	53.10	2.36	2.20

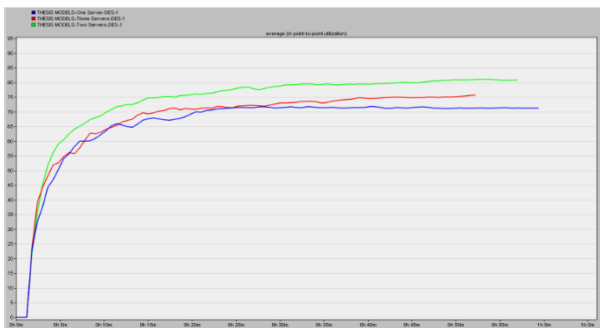


Figure 11 Average point to point utilization of router-cloud link

It was observed from Figure 11 and tables 9 and 10 that the total point to point use of cloud was better when there were three servers over the network because the cloud needed to work on the database, email, ftp, and http packets continuously. As the nodes (LAN, switches and the cloud) were imposed, some authentication policies and delays were observed. This resulted in packet diversions and cloud utilization was reduced. In the third scenario where the web, ftp and email traffics are routed, synchronized and delayed, the full utilization of the cloud network was observed as shown in Figure 11. As the traffics are designated, the servers had adequate time to process the database packets and the total utilization is reduced improving server performance. Thus at the end of analysis, it could be estimated that the total utilization of the cloud can be optimized when less servers are in the network.

5. CONCLUSIONS, FINDINGS AND RECOMMENDATIONS

Based on the findings, the following conclusions were drawn:

- ✓ The general analysis of the results revealed that, the proposed server model was well utilized for improving the user applications.
- ✓ The utilizations demonstrated that, the total performance of database application was improved when the web traffic was given a separate server. Again when there was heavy browsing across the network, the total cloud

utilization was increased and thus the server performance was low and vice versa.

- ✓ The general conclusion was that, server performance and network performance are inversely related, which implies that introducing more servers on the network correlates to increase in the network utilization performance but degrades server performance due to packet delays.
- ✓ The overall analysis was that, using one server had an initial low value but a higher value at end of the evaluation than using two or more servers. Using two or more servers had an initial higher value due to routing protocols of applications to destination servers coupled with synchronization protocols of the additional server node on the network but a lower value at the end of simulation than using a single server.
- ✓ Again, when one server in the network goes down, the other servers in the network can serve customers without a shutdown until the faulty server is restored. Samples of Average performance showed a clear distinction of the server performances.
- ✓ It also came to light that, using more than one server in a company's network had an initial high cost due to purchasing devices and installation overheads but has an overall cost effectiveness, operational efficiency and user performance than using a single server.
- ✓ Nonetheless, at a point where the servers or nodes are beyond more than required on the network, there would be the possibility of poor system performance. This is because of transmission and routing protocol complexities in processing of user requests by the servers causing delays, congestions and consequent packet dropping.

RECOMMENDATIONS

By the findings in this research, it we recommend that companies wishing to test their server performance, efficiency and reliability should consider using a virtual simulator. This is because apart from mitigating foreseeable installation problems and avoiding theft of physical components, it is also cheap and has an operational efficiency compared to using real machines.

The following are recommended for feature studies;

- ✓ A number of applications other than database, e-mail, http and ftp should be used to evaluate the performance of the server models which this paper could not cover.
- ✓ Anyone trying to embark on a similar project should give attention to the network security such as firewalls for which this research failed to highlight.
- ✓ The graphical analysis measurements should be done in 10mins and 50mins to give vivid accounts at all times to be able to capture for variations in values due to the rise and fall at certain times which the table measurements fail to do.

6. REFERENCES

1. Arora, I. & Gupta, A., 2012. Opportunities, Concerns and Challenges in the Adoption of Cloud Storage. *International Journal of Computer Science and Information Technologies*, 3(3), pp. 4543-4548.
2. Baburajan, R., 2011. The Rising Cloud Storage Market Opportunity Strengthens Vendors, infoTECH. [Online] Available at: <http://it.tmcnet.com/> [Accessed 05 May 2015].
3. Desai, M., 2012. High Performance Computing and Virtualization. *CSCi-555 Advanced Operating Systems*, pp. 1-11.
4. Kaufman, L., 2009. Data security in the world of cloud computing. *IEEE Security & Privacy*, 7(4).
5. Koponen, T. A. K. B. P. C. M. C. A. et al., 2014. 2014, April. Network Virtualization in Multi-tenant Datacentres. In *NSDI* (pp. 203-216).. *USENIX Symposium on Networked Systems (NSDI '14)*, Volume XI, pp. 203-216.
6. Luis, F., Lages, L., Abrantes, J. & Raquel, L. C., 2008. The STRATADAPT scale: A measure of marketing strategy adaptation to international business markets. *International Marketing Review* , 25(5), pp. 584-600.
7. Mell, P. & Tim, G., 2009. The NIST definition of cloud computing. *National Institute of Standards and Technology* , 53(6), p. 50.
8. Saunders, M., Lewis, P. & Thornhill, A., 2009. *Research methods for Business Students*. Uk: Pearson Professional Limited.
9. Sen, J., 2013. Security and privacy issues in cloud computing. *Architectures and Protocols for Secure Information Technology Infrastructures*, pp. 1-45.
10. Turban, E., King, D., Lee, J. & Viehland, D., 2008. Chapter 19: Building E-Commerce Applications and Infrastructure. In: *Electronic Commerce A Managerial Perspective*. s.l.:Prentice-Hall, p. 27.

Smallest Set for Reverse Selection Queries to Satisfy All

Muhammed Miah

Department of Computer Information Systems
Southern University at New Orleans
New Orleans, LA 70126, USA

Abstract: In recent years, research has moved from traditional query processing (e.g., Selection, Nearest Neighbor (NN), Top- k , Skyline), to reverse query processing (e.g., Reverse NN, Reverse Top- k , Reverse Skyline), to maximal reverse query processing (e.g., find spatial points that maximize the number of Reverse NNs), and so on. This paper considers the *Smallest Set Reverse Selection Queries Problem* also known as the *Multiple Tuple Design Problem*: Given a set of selection queries with conjunctive conditions, where the task is to create the smallest set of tuples such that each query returns at least one of these tuples. The problem is an interesting variant of the Maximal Reverse Selection Queries Problem (also referred to as the *Tuple Design Problem*) introduced by Miah et al. (2016). The paper shows that the problem is NP-Complete and develops approximation algorithms with provable approximation guarantees, as well as carefully designed heuristics that work well in practice. The paper also designs efficient exact algorithm that are feasible for moderate instances. It provides extensive experiments that demonstrate the effectiveness of the proposed algorithms.

Keywords: smallest set; reverse selection queries; multiple tuple/product design; maximize visibility, satisfy all.

1. INTRODUCTION

Significant research has been done in the area of traditional query processing, where a query and a database (set of tuples) are given, and the task is to return all tuples in the database that satisfy the query. Different query models have been developed for query processing including “selection queries” to find the set of tuples that satisfy a selection condition, “ k NN queries” to find the set of k tuples that are closest to a query tuple, “skyline queries” to find all tuples that are not dominated by any other tuple), “top- k queries” to find the top- k tuples, and so on.

Research has moved in recent years to the complementary area of reverse query processing, where in the database (set of tuples), a query log, and a potential tuple are given, and the task is to find all queries in the query log that return the given tuple. The reverse query processing also has been studied for a variety of query models, such as “reverse k NN”, “reverse top- k ”, “reverse skyline”, and so on. While traditional query processing applications focus on the customers or end users, reverse query processing applications focus on the manufacturers or sellers, e.g., helping manufacturer identify products that are most preferred by customers.

Maximal reverse query processing has been studied as well for skyline (Li et al. 2007; Li et al. 2006) and k NN (Cabello et al. 2005; Wong et al. 2009) queries with numeric attributes.

Most recently, an interesting related area of maximal reverse selection queries problem has been introduced. Given a database (set of tuples) and a query log, the task is to construct a new tuple such that the set of queries in the query log that return the new tuple is larger than for any other tuple in the database. This focused on the maximal reverse queries problem on the important class of “selection queries” over Boolean databases. Assume that a Boolean tuple (e.g., a new product) needs to be designed by selecting a subset of Boolean features (or attributes) from a large set of possible features. Assume that we are given a set of user preferences in the form of a query log (or workload) of user queries, where each query is a conjunction of positive or negative preferences

for some of the features (e.g., “Select * from Database where $a_1=0$ and $a_4=1$ and $a_6=1$ ”). The problem also known as *Tuple Design (TD) or Single Tuple Design (STD) Problem* (Miah et al. 2016).

This paper focus on an interesting variant of the Single Tuple Design problem. Instead of just designing a single tuple, we may be interested in creating a minimum number of tuples that collectively satisfy all queries in the query log. This is referred to as the *Multiple Tuple Design (MTD) problem*.

MTD (Multiple Tuple Design) Problem: Given a query log Q consisting of conjunctive selection queries over Boolean attributes, construct the smallest set T of Boolean tuples, such that for each query q in Q , there exists a tuple t in T that is returned.

The *MTD* problem has several potential applications. Consider a travel agency that wishes to design vacation packages, given the travel preferences of its clients. For example, a vacation package to Costa Rica can include some of the following attractions: beaches such as Puerto Viejo, Jaco, Flamingo, etc.; mountains and national parks such as Arenal area, Monteverde, Tortuguero, etc. The clients of the agency provide their preferences by specifying “yes”, “no”, or “don’t care” for each attraction. The travel agency might want to create a minimum set of vacation packages to satisfy all its customers, because each package induces fixed overheads to the agency, e.g., requires a dedicated vacation guide, or a transportation vehicle, etc. Likewise, a product manufacturer may wish to design and manufacture a small range of products that will cover the preferences of all customers.

One of the practical challenges is to ensure the availability of a large and rich query log to make the design of the new product/package truly effective for the customer base. While customer preferences can of course be explicitly collected through tools such as surveys and questionnaires, a very effective alternative is to implicitly collect such preferences by observing and recording user behavior on the internet – e.g., their browsing and navigation patterns on a product manufacturer or e-tailer’s website, such as the pages and

products they click on. The vast use of the Internet nowadays allows enormous amounts of such preferences to be very easily collected. However, in this paper, we do not focus on how exactly such query logs can be collected. Rather we assume the presence of such logs, and focus more on the technical challenges, i.e., designing effective and scalable solutions of the maximal reverse selection query problems (Miah et al. 2016).

The *MTD* problem is technically challenging because the problems turn out to be NP-complete (proof of NP-completeness shown later in the paper). It is therefore necessary to design good approximation algorithm that has provable approximation bound or heuristics that work well in practice, or even exact algorithms that work well for moderate problem instances. However, it is not easy to simply reuse well-known approximation algorithms for related NP-complete problems, as none of these algorithms are an exact fit for *MTD*.

Major Contributions:

- The paper considers the problem of Maximal Reverse Selection Queries for Boolean databases, and focus on specific problem, the Multiple Tuple Design (*MTD*) problem. The first result is to show that *MTD* is NP-complete. The proof for *MTD* uses a reduction from Graph Coloring (*GC*).
- Approximation algorithms for the *MTD* problem are developed. The first algorithm has a provable approximation bound, and is based on a combination of known approximation algorithms (and their approximation factors) of two separate NP-complete problems. However, this algorithm is mainly of theoretical interest as its approximation factor is quite large.
- Two other more practical approximation algorithms are also developed. While these algorithms do not guarantee any approximation bounds, they are scalable and are shown to have small approximation factors in practice.
- The extension of the problem to the Categorical and Numeric Databases also discussed.
- Detailed performance evaluations are performed on real and synthetic data to demonstrate the effectiveness of the developed algorithms.

The rest of the paper is organized as follows. Section 2 provides formal problem definitions for *MTD*. Section 3 analyzes the computational complexities for the problems. Section 4 presents approximation algorithms with provable bounds. Section 5 provides approximation heuristics that are scalable and have small approximation factors in practice. Section 6 presents the result of extensive experiments. Section 7 provides the extension of the problems to other databases. Related work is discussed in Section 8, conclusions in Section 9, and Section 10 provide references.

2. PROBLEM FRAMEWORK

To define the problem more formally, we need to develop a few abstractions.

Attributes: Let $A = \{a_1 \dots a_M\}$ be a set of Boolean attributes (or elements, or features).

Query (with negation): We view each user query as a subset of attributes and/or negation of attributes. The semantics is conjunctive, e.g., query $\{a_1, a_3\}$ is equivalent to " $a_1 = 1$ and a_3

$= 1$ ". We also consider queries with negations, e.g., $\{a_1, \sim a_2\}$ is equivalent to " $a_1 = 1$ and $a_2 = 0$ ". The remaining attributes for which values are not mentioned in the query are assumed to be "don't care", i.e., the value in the new designed tuple can be either 0 or 1.

Query Log or Workload: Let $Q = \{q_1 \dots q_S\}$ be a collection of queries.

For the *MTD* problem, we need to design a number of tuples. In fact, we need to design the minimum number of tuples such that together they satisfy all the queries in the query log. Below is more formal definitions of the problems introduced in Section 1:

Multiple Tuple Design (*MTD*) Problem: Given a query log Q with conjunctive semantics where a query can have negations, design the minimum number of tuples (assign value $[0, 1]$ for each attribute in each tuple) such that for each query of Q there exists a tuple that satisfies it.

Example 1. Consider Table 1 which shows a query log for a vacation package application, containing $S=6$ queries and $M=6$ attributes where each tuple (query) represents the preferences of a user. A query has values 1, 0, or ?, where 1 means the attribute must be present, 0 means the attribute must not be present, and "?" means "don't care". As we can see for this example that if we design three new packages as t_1 (with *Beach* = 1, *Boating* = 0, *Casino* = 0, *Fishing* = 1, *Historical Site* = 1, *Museum* = 0 which satisfies three queries q_2, q_4 and q_6), t_2 (with *Beach* = 0, *Boating* = 1, *Casino* = 1, *Fishing* = 0, *Historical Site* = 1, *Museum* = 1 which satisfies two queries q_3 and q_5), and t_3 (with *Beach* = 1, *Boating* = 0, *Casino* = 1, *Fishing* = 1, *Historical Site* = 0, *Museum* = 0 which satisfies one query q_1), we can satisfy all 6 queries in Table 1. No other combination of three or less packages will satisfy all queries. □

Table 1. Query Log Q for Running Example

Query ID	Beach	Boating	Casino	Fishing	Historical Site	Museum
q_1	1	0	1	?	?	?
q_2	1	?	0	?	1	?
q_3	0	?	1	?	1	?
q_4	?	0	0	1	1	?
q_5	?	1	?	0	?	1
q_6	1	0	0	?	?	0

3. COMPLEXITY RESULTS

The *MTD* problem is NP-complete.

Theorem 1: *MTD* is NP-complete.

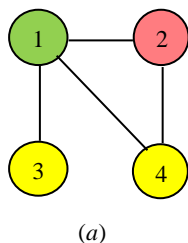
Proof: The problem is clearly in NP as a proposed solution can be easily verified in polynomial time by a single pass over the query log. To prove that it is NP-complete, we reduce the Graph Coloring (*GC*) problem to *MTD*. In the *GC* problem, we wish to assign colors to the vertices of a graph such that no two adjacent vertices share the same color, and are required to

find the minimum number of colors needed to color the graph (this number is known as the chromatic number of the graph).

The reduction is as follows. Let $G = (V, E)$ be the graph in the GC instance. Edges will correspond to Boolean attributes and vertices will correspond to queries. Each edge $e_i = (v_j, v_k) \in E$ represents a Boolean attribute a_i , with the condition $a_i = 1$ a part of the query q_j corresponding to vertex v_j and the condition $a_i = 0$ a part of the query q_k corresponding to vertex v_k . (i.e., with one end point representing the positive literal a_i and the other end point representing the negative literal $\sim a_i$). To make it little more clear, when considering vertex v_j (for query q_j), then only considering the end point of edge e_j connected to v_j , not the end connected to v_k ; and similarly when considering vertex v_k (for query q_k), then the same edge e_j is being represented as e_k only considering the end point of edge e_k connected to v_k , not the end connected to v_j . Thus, each vertex v_j represents a query q_j , where the query is represented as a conjunction of the literals corresponding to the end points of all edges that are incident to v_j , and the remaining attributes are “don’t care”s.

Thus, Figure 1(a) shows a 3-colorable graph G with a valid coloring and Figure 1(b) shows the corresponding query log for the MTD instance. It is easy to see that finding the chromatic number of G is equivalent to finding the smallest set of tuples (with an assignment of $\{0, 1\}$ values to each Boolean attribute) in the MTD instance that satisfies all the queries. \square

The GC problem is NP-hard only for graphs with chromatic number > 2 and hence the MTD problem is NP-hard only if we need to design more than 2 tuples to cover all queries.



(a)

Query ID	$a_{1,2}(a_{2,1})$	$a_{1,3}(a_{3,1})$	$a_{1,4}(a_{4,1})$	$a_{2,4}(a_{4,2})$
q_1	1	1	1	0
q_2	1	0	0	1
q_3	0	1	0	0
q_4	0	0	1	1

(b)

Figure 1. An instance graph and corresponding query log.

4. APPROXIMATION ALGORITHM WITH PROVABLE BOUND

Although the *Graph Coloring* (GC) problem is used to prove NP-completeness of MTD , GC does not have a good bounded-

factor approximation algorithm. In fact, the best known approximation algorithm has an approximation factor that can be very close to the number of vertices in the graph. Hence such algorithms are not useful for developing bounded-factor approximation algorithms for MTD . Moreover, an approximation algorithm for GC is not directly applicable to MTD problem.

Instead, proposed approach for developing an approximation algorithm for MTD is based on two steps. We first define a problem that is more convenient than MTD for developing approximation algorithms, called the *Multiple Tuple Selection* problem (MTS). We then consider an approximation algorithm for MTS , and make modifications so that it can eventually be used to solve MTD . The modifications entail making repeated invocations to the approximation algorithm for STD (Miah et al. 2016) for different instances of STD . The details are given below.

The *Multiple Tuple Selection* (MTS) problem is almost the same as the MTD , except that in addition to the query log, a set of candidates tuples P is also given, and we are restricted to select the minimum subset of tuples T from this set P .

Lemma 1: MTS is NP-Complete

Proof: The problem is clearly in NP as a solution can be verified in polynomial time. To prove NP-completeness, we reduce from the *Hitting Set* (HS) problem (Garey and Johnson 1979). Given a ground set Z of elements, and a collection Y of subsets of Z , the goal of the HS problem is to find the smallest subset $H \subseteq Z$ of elements that hits every set of Y .

The reduction of HS to MTS is as follows. Assume MTS has V candidate tuples, S queries, and M attributes; and HS has n sets and m elements. For each element e_i in HS , create a query q_i and an attribute a_i in MTS . Set $a_i(q_i) = 1$ and the rest of the attributes of q_i to “don’t care”. For each set s_i in HS , create a tuple t_i in MTS . For each element e_j in s_i , set $a_j(t_i) = 1$. Set the rest of the attributes of t_i to 0. Then, a solution to HS is a solution to MTS and vice-versa.

Note that in the above reduction, we set $S = M = m$ and $V = n$. That is, we use as many attributes as queries. If we would assume that $M = \log(S)$, then the above reduction is not valid and it is an open problem whether the problem is NP-complete or not. \square

Even though the MTS problem is NP-complete (as is MTD), MTS has an advantage over MTD - it can be solved using the well-known greedy approximation algorithm for the *SETCOVER* problem (Cormen et al. 2001), which we describe below.

An instance (X, F) of the *SETCOVER* problem consists of a finite set X and a family F of subsets of X , such that every element of X belongs to at least one subset of F . The goal is to find a minimum subset $E \in F$ whose members cover all of X . A greedy approximation algorithm (Cormen et al. 2001) provides an approximation bound of $H(\max \{|B| : B \in F\})$ where $H(d)$ represents the d^{th} harmonic number which is equal to $\log(\max \{|B| : B \in F\}) + O(1)$. The greedy algorithm works as follows: (i) at each iteration, the algorithm picks the set with highest number of elements from the sets not picked yet, and (ii) the process repeats until all elements are covered.

We can directly relate MTS to the *SETCOVER* problem as follows: let us assume X is the query log (Q) and F is the set of given tuples (P), where each package represents the set of queries that are satisfied by it. Now the goal is to find the minimum set of tuples T from P such that they cover all the

queries in Q . Thus MTS also has the same approximation bound as the $SETCOVER$ problem which would be $\log(\max\{|V| : V \in P\}) + O(1)$, where V is a subset $V \in P$ which covers all queries in Q . Thus, in the worst case the approximation factor for MTS is $\log(S)$ where S is the number of queries in the query log Q .

Extending this approach to the MTD problem is challenging as we do not have a set of candidate packages. Moreover, it is not possible to simply first enumerate all possible packages, as this is exponential in the number of attributes! However, one of the interesting contributions of this paper is the observation that *we can avoid this enumeration by combining the approximation algorithm for STD (discussed earlier) with the greedy approximation algorithm for $SETCOVER$ to eventually design an approximation algorithm for MTD* . Essentially, we take the greedy algorithm for $SETCOVER$, and make modifications that entail making repeated invocations to the approximation algorithm for STD for different STD instances.

This combined algorithm proceeds as follows: (i) at each iteration the algorithm makes a call to the approximation algorithm for STD (Miah et al. 2016) over the not-yet-satisfied queries, which returns a tuple such that the number of new queries that are satisfied is at least $ck/2^k$ times the number of new queries that would have been satisfied by an optimal tuple, and (ii) the process repeats until all queries are satisfied.

The following theorem shows that the above algorithm has a provable approximation bound.

Theorem 2: *The number of tuples returned by the approximation algorithm described above is at most $(2^k/ck)\log(S)$ times the number of tuples returned by an exact algorithm for MTD .*

Proof: Let T_{min} be the optimal (minimum) set of tuples that together satisfy all queries in the query log Q . Let $T' = t'_1, t'_2, \dots, t'_x$ be the sequence of tuples returned by the above approximation algorithm. We shall show that $x=|T'| \leq |T_{min}|(2^k/ck)\log(S)$.

Let us imagine that each tuples is allotted a weight of 1, and the weight is evenly distributed to all queries that are satisfied for the first time by that package. For example, if t'_1 satisfies queries q_1 and q_2 , then $wt(q_1) = wt(q_2) = 1/2$; if tuple t'_2 satisfies query q_3, q_4 , and q_5 , then $wt(q_3) = wt(q_4) = wt(q_5) = 1/3$. It is easy to see that the accumulated weight of all the queries aggregate to x .

Consider any tuple t_i of T_{min} . Let $t'_{i,1}, t'_{i,2}, \dots, t'_{i,r}$ be the order in which the approximation algorithm returns tuples that have non-empty intersection with t_i , i.e., that there exists at least one query in the query log that is satisfied by both $t'_{i,j}$ and t_i .

To make notation convenient, we will refer to any tuple t as the set of queries that it satisfies, and $|t|$ as size of this set.

Let $u_{i,j} = |t' - (t'_{i,1} \cup t'_{i,2}, \dots \cup t'_{i,j})|$. Thus, of all the queries that are satisfied by t' , $u_{i,j}$ refers to the number of queries that are still not satisfied by the approximation algorithm at the time the last generated tuple is $t'_{i,j}$.

Consider $t'_{i,j+1}$. We argue that $|t'_{i,j+1}| \geq (ck/2^k) u_{i,j}$. Because if this were not so, then the greedy approximation algorithm for STD would not have returned $t'_{i,j+1}$, since the number of remaining queries it would have satisfied would have been too small and its known approximation bound would have been violated – in fact, t_i itself would have been a better package to return next instead of $t'_{i,j+1}$.

Thus, the aggregate weight of all queries satisfied by t_i is

$$\leq \sum_{1 \leq j \leq r} (u_{i,j+1} - u_{i,j}) \frac{1}{\left(\frac{ck}{2^k}\right)^{u_{i,j}}} \leq \sum_{1 \leq j \leq r} \frac{1}{\left(\frac{ck}{2^k}\right)^j} \leq \left(\frac{2^k}{ck}\right) \log(S)$$

Thus, the aggregate weight of all queries is

$$x = |T'| \leq |T_{min}| \left(\frac{2^k}{ck}\right) \log(S)$$

□

In summary, the above discussions demonstrate the existence of bounded approximation factor algorithms for the MTD problem. However, while interesting from a conceptual point of view, in practice these approximation factors are too suboptimal to be useful, and the algorithms themselves require complex in-memory implementations based on semi-definite programming (SDP) relaxation methods. More practical approximation algorithms are needed that work well for moderate as well as large problem instances. Such algorithms are discussed in the next section.

5. SCALABLE APPROXIMATE ALGORITHMS

The approximation algorithm of Section 4 is useful for theoretical purposes, but it is not very feasible in practice. This is because that is main memory algorithms based on semi-definite programming (SDP) relaxation, which do not scale for large instances. Moreover, the bounded approximation factors are too suboptimal to be useful in practice. This section proposes scalable approximation algorithm MTD , which are shown to perform well in terms of scalability and approximation error in Section 6.

The solution for MTD involves a combination of the solution of multiple instances of STD . Miah et al. (2016) proposed an exact algorithm for STD based on Signature Tree data structures. A depth first creation of the signature tree is adapted, where at each downward step it picks an attribute that will be used as the splitting node. Further, for every traversed node v , it checks if this node has a chance to be the best package. That is, it compares the number of satisfied queries in v to the *current maximum number C of satisfied queries*. If v fails this test, it prunes the whole subtree rooted at v , since the children of v satisfy at most as many queries as v . During the depth-first traversal, it first traverses the child that contains the maximum number of queries, since the objective is to create an assignment of values to all attributes that maximizes the number of satisfied queries. When it finds the node v with the maximum C , it creates the corresponding “best” tuple by traversing the path from v to the root and assigning attribute values accordingly.

Figure 2 shows the pseudocode of *SigTreeSTD* exact algorithm for STD (Miah et al. 2016)

Algorithm: SigTreeSTD
Signature Tree T //initially single node u that contains all queries in Q , that is $u=Q$.
Let C be the current max # queries in a leaf node of the signature tree. Initially $C \leftarrow 0$
Stack V
 $V.push(T)$ // T is root node of T
While V not empty do
 $u \leftarrow V.top()$
 If both children of u have been processed before or there is no splitting attribute then
 update C //if $|u| > C$ then $C = |u|$
 $V.pop()$;
 Continue;
 Find next splitting attribute A for u
 If $|u_{A=0}| > |u_{A=1}|$ and $u_{A=0}$ not processed before // $u_{A=1}$ is the set of queries from u that satisfy $A=1$ (similar for $u_{A=0}$)
 $V.push(u_{A=0})$ //dfs to create tree for left child.
 Else if $u_{A=1}$ not processed before
 $V.push(u_{A=1})$ //dfs to create tree for right child
Return assignment for C

Fig 2. Pseudocode of SigTreeSTD exact algorithm for STD (Miah et al. 2016)

Miah et al. (2016) also proposed an approximation algorithm *HeuristicSTD* for STD, the pseudocode is shown in Figure 3.

Approx Algorithm: HeuristicSTD
Let Q be the query log, $A (a_1 \dots a_M)$ be the attributes in Q
Complement the query log ($\sim Q$) // convert 1 to 0 and 0 to 1, also convert conjunctive form to disjunctive form
For (int $i = 1$ to M)
 If $\sim Q$ not empty
 Count # of queries satisfied both for $a_i = 1$ and $a_i = 0$.
 Assign the value of a_i that gives the minimum count
 Remove queries from $\sim Q$ satisfied by the value of a_i
Return the attributes assignment

Fig 3. Pseudocode of HeuristicSTD: approximation algorithm for STD (Miah et al. 2016)

We consider two variants for STD: one using the exact or optimal STD algorithm, and the other using the heuristic of STD (Miah et al. 2016). In particular, the algorithms are as follows:

- Apply algorithm for STD and add solution assignment (tuple) to result.
- Remove the queries from the query log that are satisfied by assignment in step (a).
- Repeat steps (a) and (b) until no further queries or attributes are left.

Figure 4 displays the pseudo-code of two approximation algorithms for MTD. Note that the way that the idea of

greedily removing satisfied queries is inspired by a SETCOVER heuristic (Cormen et al. 2001).

The SigTreeMTD algorithm has approximation bound $\log(S)$, where S is the number of queries in Q , since the STD component of the algorithm is exact. There is no approximation bound for HeuristicMTD, but it is shown to perform well in Section 6.

Algorithm: SigTreeMTD
While Q not empty do // Q is the query log
 Apply algorithm SigtreeSTD on Q
 Remove queries from Q those satisfy the assignment

Algorithm: HeuristicMTD
While Q not empty do
 Apply algorithm HeuristicSTD on Q
 Remove queries from Q those satisfy the assignment

Fig 24. Pseudocode of SigTreeMTD and HeuristicMTD approximation algorithms for MTD

6. EXPERIMENTS

The main performance indicators are (a) the time cost of exact and approximation algorithms, and (b) the approximation quality of approximation algorithms.

6.1 System Configuration and Datasets

System Configuration: We used Microsoft SQL Server 2000 RDBMS on a P4 3.2-GHZ PC with 1 GB of RAM and 100 GB HDD for our experiments. Algorithms are implemented in C#.

Datasets: Datasets of products and product queries are used. Note that products are just one of the possible instantiations of the more general packages of this paper. We used real and synthetic datasets (query logs). In specific, we use two datasets: (i) REAL: real query log, and (ii) REAL+: synthetic query log generated from the real query log.

Real query log (REAL): 237 queries collected for cell phones from university users and friends through an online survey. The survey was designed with 30 Boolean features such as Bluetooth, Wi-Fi, Camera, Speakerphone and so on. Users were asked to select the features they prefer to have (positive) and most likely not to have (negative) in their cell phones. Users selected 3-6 positive and 1-2 negative features on average. Hard disk was a popular negative feature.

Synthetic query log generated from real query log (REAL+): As the real query log is very small, it is inappropriate for scalability experiments. So larger datasets were generated from the real query log. A total of 251,575 queries were generated as follows: at each step we randomly select a query from the REAL query log, randomly select two of its attributes and swap their values. We also generate datasets for a fixed size of query log for varying number of attributes (10, 15, 20, 25, 30).

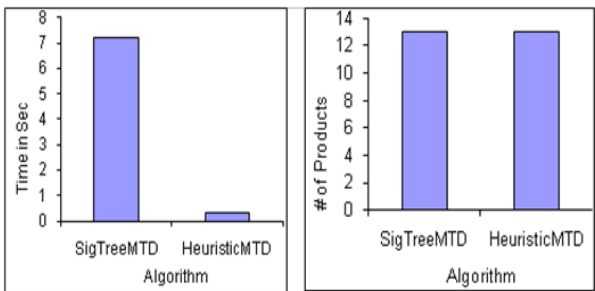
Table 2 summarizes the query logs or datasets.

Table 2. Summary of Query Logs (Datasets)

Query log	# of attributes	Query log size
REAL	30	237
REAL+_30	30	25K, 50K, ..., 200K
REAL+_1000	10, 15, ..., 30	1000

6.2 Experimental Results

Figure 5 shows the performance and quality of the algorithms for the REAL dataset. Figures 6 and 7 show the performance of the algorithms for varying query log size and number of attributes respectively, for REAL+ dataset. As we can see from the graphs, the *HeuristicMTD* algorithm is much more efficient than the *SigTreeMTD* which is developed based on exact algorithm of *STD*. The missing data in Figure 6 is due to the very slow speed of this algorithm for large datasets. The running time of *SigTreeMTD* algorithm increases exponentially as the number of total attributes increases.



(a) Time cost (b) Quality

Fig 5. Time cost and Quality for REAL dataset

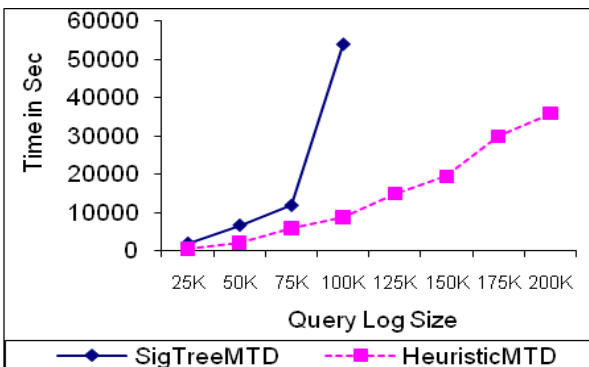


Fig 6. Time cost for varying query log size for REAL+_30

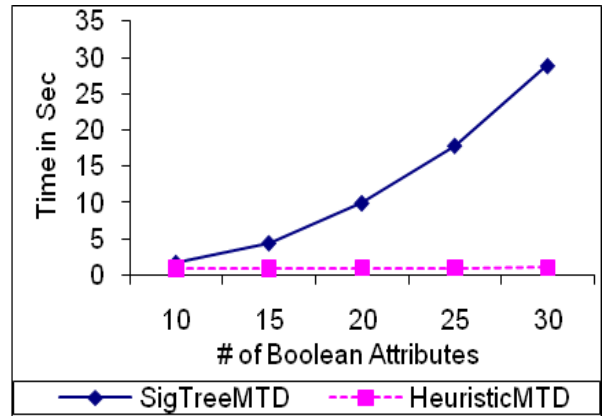


Fig 7. Time cost for varying # of attributes for REAL+_1000

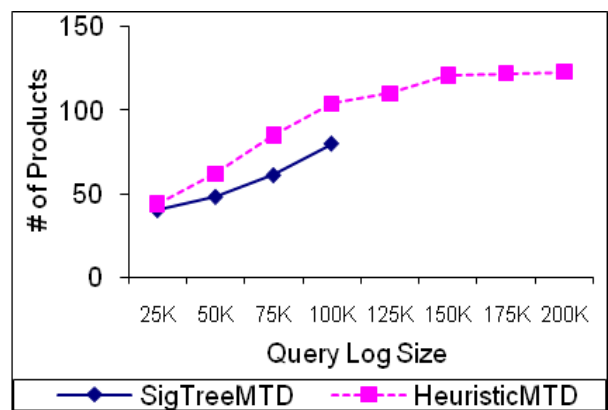


Fig 8. Quality for varying query log size for REAL+_30

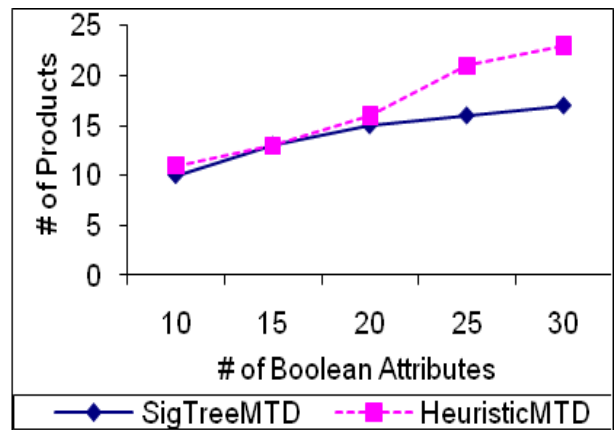


Fig 9. Quality for varying # of attributes for REAL+_1000

Figures 8 and 9 show the quality (number of products need to design to satisfy all queries in the query log) of the approximation algorithm for varying query log size and number of attributes respectively for REAL+. For the same reason as in Figure 6, Figure 9 also has missing data for *SigTreeMTD* algorithm.

7. EXTENSION TO OTHER TYPES OF DATABASES

The Boolean problems discussed above can be extended to categorical and numerical databases as well.

7.1 Problem Framework

Categorical: Can be converted into a set of Boolean attributes according to its distinct categorical values.

Numeric: Comprises of a subset of attributes taking numeric values, while the remaining attributes for which values are not specified are assumed to be “don’t care” (represented as “?”), i.e., in the newly designed tuple the numeric attribute can have any value between the range of values it can take. For e.g, query $\{a_1=10, a_3=5.2\}$ is equivalent to “ $a_1=10$ and $a_3=5.2$ ”, while the remaining attributes are ?.

7.2 Complexity Results

Corollary 1: *MTD* is NP-complete for categorical databases. It is easy to see that the reduction from *GC* to the Boolean *MTD* instance can be readily extended to reduce *GC* to the categorical *MTD* instance by having as many edges as the number of distinct categorical values an attribute can take.

Corollary 2: *MTD* is NP-complete for numeric databases. The proof of NP-completeness for numeric databases follows from Corollary 1 where the distinct categorical attribute values in the graph edges are replaced by numeric values from the range of values that numeric attribute can have.

7.3 Numeric SigTreeMTD Algorithm

The Boolean version of the algorithm a_i extended as follows. For each attribute, let $R(a_i)=r_1, \dots, r_s$ be the set of range conditions specified for a_i in the query log Q . Let $E(a_i)$ be the list of endpoints of the ranges in $R(a_i)$, in ascending order. Then, we define $I(a_i)$ as the list of intervals created from $E(a_i)$, one interval for each two consecutive endpoints. E.g. if $r_1=(2,3)$, $r_2=(3,5)$, $r_3=(4,7)$, then $E(a_i)=\{2,3,4,5,7\}$ and $I(a_i)=\{[2,3], [3,4], [4,5], [5,7]\}$. Note that $|E(a_i)| \leq 2 \times |Q|$ and $|I(a_i)| \leq 2 \times |Q| - 1$, where $|Q|$ is the number of queries in Q . Then, each attribute has $|I(a_i)|$ children, denoting all possible intervals for a_i . Each path from root to a node u can be viewed as a list of intervals, one for each attribute along the path. To compute the number of queries for u , we make a pass on Q and count the number of queries that satisfy all the conditions along the path from root to u . The rest of the algorithm and pruning conditions are the same as the Boolean version, where instead of having only two values 0,1 to choose from at each step, we choose among the intervals in $I(a_i)$.

8. RELATED WORK

The most related work to work of this paper is the the maximal reverse problem or *Single Tuple Design (STD) Problem*: Given a database (set of tuples) and a query log, the task is to construct a new tuple such that the set of queries in the query log that return the new tuple is larger than for any other tuple in the database (Miah et al. 2016). However, this paper focus on an interesting variant of the Single Tuple Design problem. Instead of just designing a single tuple, we are interested in creating a minimum number of tuples that collectively satisfy all queries in the query log. This is referred to as the *Multiple Tuple Design (MTD) problem*.

There is been recent interest in the area of Reverse Query Processing for various query models such as *kNN*, *Top-k*, *skyline*, and so on (Binning et al. 1994, 2007; Dellis and Seeger 2007; Korn and Muthukrishnan 2000; Vlachou et al. 2010). Unlike traditional query processing, the applications

are not from a customer’s point of view, but rather are from the manufacturer’s point of view, i.e., of how to determine the set of customers that find a particular product appealing. Reverse Nearest Neighbor (*RNN*) queries were first investigated by Korn and Muthukrishnan (2000). Given any query point q , Reverse NN is to determine the set $RNN(q)$ of reverse nearest neighbors. Reverse Skyline Queries (Dellis and Seeger 2007) considers for a multidimensional data set D the problem of dynamic skyline queries according to a query point q . This kind of dynamic skyline corresponds to the skyline of a transformed data space where point q becomes the origin and all points of D are represented by their distance vector to q . The reverse skyline query returns the objects whose dynamic skyline contains the query object q . Recent work on Reverse top- k queries (Vlachou et al. 2010) is from the perspective of the product manufacturer. The problem is, given a potential product, which are the user preferences for which this product is in the top- k query result?

The work in this paper is different than all the works on reverse query processing discussed above. We are not given the set of data tuples to pick from, instead we have to design a set of new tuples (*MTD*) that satisfy maximum number of queries in the given query log. In this regard, our STD problem is somewhat similar to maximal reverse query processing problems that has recently received some attention for skyline (Li et al. 2007; Li et al. 2006) and *kNN* (Cabello et al. 2005; Wong et al. 2009) queries with numeric attributes. Our work is different from these because we consider selection queries over Boolean attributes.

Works on dominant relationship (Li et al. 2006) and dominating neighborhood (Li et al. 2007) uses skyline query semantics assuming that attributes are min/max, that is, all users have the same preference for an attribute (e.g., 2 doors is always better than 4 doors). Further, they assume there is a profitability plane which simplifies the algorithm given that the optimal solution is a point on the profitability plane. In contrast, in work of this paper along with the work of Miah et al. (2016) users may have opposite preferences for the same attribute, and the algorithms can be used with or without a profitability plane.

Miah et al. (2009) tackled another related problem of maximizing the visibility of an existing object by selecting a subset of its attributes to be advertised. The main problem was: given a query log with conjunctive query semantics and a new tuple, select a subset of attributes to retain for the new tuple so that it will be retrieved by the maximum number of queries. The work did not consider negated conditions as in the work of this paper.

Optimal product design or positioning is a well-studied problem in Operations Research and Marketing. Shocker and Srinivasan (1974) first represented products and consumer preferences as points in a joint attribute space. After that, several approaches and algorithms (Albers and Brockhoff 1977; Albers and Brockhoff 1980; Albritton and McMullen 2007; Gavish et al. 1983; Gruca and Klemz 2003; Kohli and Krishnamurti 1989) have been developed to design/position a new product. Works in this domain require direct involvement (one or two step) of consumers and users are usually shown a set of existing alternative products (predesigned) to choose or set preferences. Users in this domain in fact do not get to select the attributes or features they like and do not like. Instead of involving users directly in the process of designing new products, this paper uses previous user search queries to model user preferences, since it is easy to collect the preferences (search queries) for large number of Internet users

nowadays. This paper also consider large query logs to design the new set of products and allow users to express their interests in attribute or feature level in terms of positive, negative and “don’t care”.

The *MTD* problem can be viewed as the segmentation problem (Kleinberg et al. 1998) for the *STD* problem (Miah et al. 2016). However, in *MTD* the size of each segment is not given.

9. CONCLUSIONS

This paper investigated the problem of designing smallest set of tuples for maximal reverse selection queries - given a set of selection queries with conjunctive conditions (where a query can have negations), create the smallest set of tuples that that collectively satisfy all queries in the query log. The problem has several natural applications, such as designing best vacation packages, designing new products, and so on. The paper shows the difference of the proposed problem from the existing techniques in various fields such as marketing, product design, operation research, query processing, etc. The paper considers several interesting variants of the problem as well as various types of databases such as Boolean, categorical, and numerical. It proves intractability results, and provide approximation algorithms, some of which are shown to work well in practice. A future direction is to extend the problem to develop more scalable algorithms for categorical, numeric, and possibly text data and different query semantics such as top-*k* and skyline queries.

10. REFERENCES

- [1] Albers, S., and Brockhoff, K. “A procedure for new product positioning in an attribute space”, European Journal of Operational Research, 1, 4 (Jul 1977), 230-238.
- [2] Albers, S., and Brockhoff, K. “Optimal Product Attributes in Single Choice Models”, Journal of the Operational Research Society (1980) 31, 647–655.
- [3] Albritton, D. M., and McMullen P. R. “Optimal product design using a colony of virtual ants”, European Journal of Operational Research, 176, 1 (Jan 2007), 498-520.
- [4] Cabello, S., Diaz-Banez, J. M., Langerman, S., Seara, C., and Ventura, I. 2005. “Reverse facility location problems”, Canadian Conference on Computational Geometry.
- [5] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. 2001. Introduction to Algorithms, The MIT Press.
- [6] Dellis, E., and Seeger, B. 2007. “Efficient computation of reverse skyline queries”, VLDB.
- [7] Garey, M. R., and Johnson, D. S. 1979. Computers and Intractability: A Guide to the Theory of NP-Completeness, New York, W.H. Freeman, ISBN 0-7167-1045-5.
- [8] Gavish, B., Horsky, D., and Srikanth, K. 1983. “An Approach to the Optimal Positioning of a New Product”, Management Science, 29, 11, 1277-1297.
- [9] Gruca, T. S., and Klemz, B. R. 2003. “Optimal new product positioning: A genetic algorithm approach”, European J. of Operational Research, 146, 3, 621-633.
- [10] Kleinberg, J., Papadimitriou, C., and Raghavan, P. 1998. “Segmentation Problems”, ACM Symposium on the Theory of Computing, 473-482.
- [11] Kohli, R., Krishnamurti, R. 1989. “Optimal product design using conjoint analysis: Computational complexity and algorithms”, European Journal of Operational Research, 40, 186–195.
- [12] Korn, F., Muthukrishnan, S. 2000. “Influence sets based on reverse nearest neighbor queries”, SIGMOD.
- [13] Li, C., Tung, A. K. H., Jin, W., and Ester, M. 2007. “On Dominating Your Neighborhood Profitably”, VLDB, 818-829.
- [14] Li, C., Ooi, B. C., Tung, A. K. H., Wang, S. 2006. “DADA: a Data Cube for Dominant Relationship Analysis”, SIGMOD.
- [15] Miah, M., Das, G., Hristidis, V., and Mannila, H. 2009. “Determining Attributes to Maximize Visibility of Objects”, IEEE Transactions on Knowledge and Data Engineering (TKDE) vol. 21 no. 7, pp. 959-973.
- [16] Miah, M., Omar, A. 2016. The International Academy of Business and Public Administration Disciplines (IABPAD) Conference Proceedings.
- [17] Shocker, A. D., and Shrinivasan, V. 1974. “A consumer-based methodology for the identification of new product ideas”, Management Science, 20, 6, 921-937.
- [18] Vlachou, A., Doulkeridis, C., Kotidis, Y., and Norvag, K., 2010. “Reverse Top-k Queries”, ICDE.
- [19] Wong, R. C-W., Özsu, M. T., Yu, P. S., Fu, A. W-C., and Liu, L. 2009. “Efficient method for maximizing bichromatic reverse nearest neighbor”, VLDB.

Mind Controlled Devices

Disha Moreshiya
PCB Designer
Nagpur, India

Nayan Dhoke
Programmer
Nagpur, India

Akshy Sadawarti
Presentator
Nagpur, India

Parthao Lad
Planner
Nagpur, India

Abstract: 'Mind Controlled devices' this project is totally based on the mind waves that is alpha waves. That mind waves has frequency in between 7.5Hz to 12.5Hz. All biological system is operates in same frequency range. Now this thing will have many advantages to control many different devices like AC or DC devices. This will be very use full for many fields like in defense, medical uses, industry, also in daily life. This project will increase the human limits. It also make life easy. Mind Controlled devices will consist of EEG electrodes some instrumental amplifies microcontroller boards through the instrumental amplifier we will trying to amplify the alpha signals which is read able to the microcontroller board.

Keywords: - Arduino, instrumental amplifier (INA), L293D, Controlling unit, alpha waves, mind reading, TTL Logic converter.

1. INTRODUCTION

Alpha waves are neural oscillations in the frequency range of 7.5–12.5 Hz arising from synchronous and coherent (in phase or constructive) electrical activity of thalamic pacemaker cells in humans. They are also called Berger's wave in memory of the founder of EEG. Alpha waves are one type of brain waves detected either by electroencephalography (EEG) or magneto encephalography (MEG) and predominantly originate from the occipital lobe during wakeful relaxation with closed eyes. Alpha waves are reduced with open eyes, drowsiness and sleep. Historically, they were thought to represent the activity of the visual cortex in an idle state. More recent papers have argued that they inhibit areas of the cortex not in use, or alternatively that they play an active role in network coordination and communication. Occipital alpha waves during periods of eyes closed are the strongest EEG brain signals.

In this Project we are using Arduino, L293D, INA332, Relay, Other things. In this we are amplifying the brain waves through INA332 then that INA 332 give the signal to Arduino now this board utilizing and identifying the purpose of the signal and converting in to movement using motor relays or any other AC or DC Devices. The brain waves will read through the EEG Electrode (Electroencephalogram).

History

- 1) Vladimirovich (1912) He described the human alpha and beta rhythms
First animal EEG study (dog)
1929: Hans Berger developed electroencephalography, the graphic representation of the difference in voltage between two different cerebral locations plotted over time
Cybulski (1914)
Cybulski acquired camera, which he use to perform the first electroencephalogram of the brain wave monkey

- 2) Berger (1924)

First human EEG recordings

'Invented' the term electroencephalogram (EEG)

- 3) American EEG Society formed in 1947

American Clinical Neurophysiology Society (ACNS) was founded in 1946, originally named the American Electroencephalographic Society (AEEGS). In 1995, AEEGS changed its name to the current designation, ACNS, to more accurately encompass the scope of its mission and goals. ACNS is the major professional organization in the United States devoted to the establishment and maintenance of standards of professional excellence in clinical neurophysiology in the practice of neurology, neurosurgery and psychiatry.

- 4) Aserinsky & Kleitman (1953)

First EEG recordings of REM sleep (.75"). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

Quantitative EEG (QEEG) is digital recording of the EEG. For decades it was only possible to record the various brain waves on paper with the traditional polygraph. The EEG rhythms were amplified and used to drive pens, one for each recording electrode. As the pens fluctuated from the EEG rhythms, a long piece of graph paper was dragged under the pens by a motor, creating the graph of the electrical activity on the outer surface of the brain.

Historically, they were thought to represent the activity of the visual cortex in an idle state. More recent papers have argued that they inhibit areas of the cortex not in use, or alternatively that they play an active role in network coordination and communication. Occipital alpha waves during periods of eyes closed are the strongest EEG brain signals Neural oscillations in the frequency range of 7.5–12.5 Hz arising from synchronous and coherent (in phase or constructive) electrical activity of thalamic pacemaker cells in humans

Components

1) TI Parts

i) INA332



Figure 1 INA332

The INA332 and INA2332 are rail-to-rail output, low-power CMOS instrumentation amplifiers that offer wide range, single supply, and bipolar-supply operation. Using a special manufacturing flow, the INA332 family provides the lowest cost available, while still achieving low-noise amplification of differential signals with low quiescent current of $415\mu\text{A}$ (dropping to $0.01\mu\text{A}$ when shut down). Returning to normal operation within microseconds, this INA can be used for battery or multichannel applications.

ii) MSP430F5529 USB Experimenter's Board

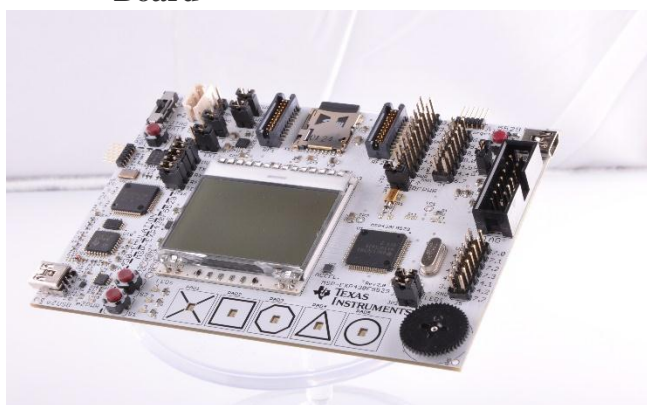


Figure 2 MSP430F5529 USB Experimenter's Board

The MSP-EXP430F5529 Experimenter Board is a development platform based on the MSP430F5529 with integrated USB. The Experimenter Board showcases the abilities of the latest family of MSP430s and is perfect for learning and developing USB-based applications using the MSP430. The features include a 102x64 dot-matrix LCD, micro SD memory card interface, 3-axis accelerometer, five capacitive-touch pads, RF EVM expansion headers, nine LEDs, an analog thumb-wheel, easy access to spare F5529

pins, integrated Spy-Bi-Wire flash emulation module, and standard full JTAG pin access. The kit is pre-programmed with an out-of-box demo to immediately demonstrate the capabilities of the MSP430 and Experimenter Board. This document details the hardware, its use, and the example software.

This equipment is intended for use in a laboratory test environment only environments may cause interference with radio communications, in which case the user, at his own expense, will be required to take whatever measures may be required to correct this interference.

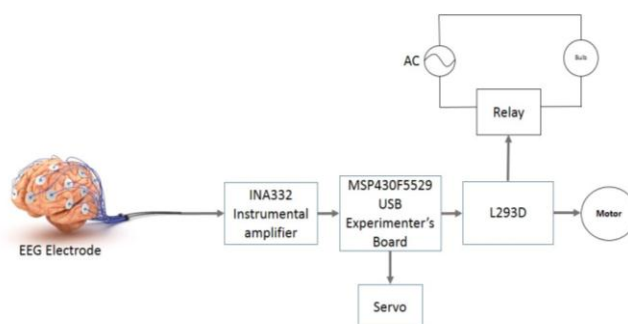
iii) L293D



Figure 3 L293D Motor driver IC

1.1 The L293 and L293D are quadruple high-current half-H drivers. These devices are designed to drive a wide array of inductive loads such as relays, solenoids, DC and bipolar stepping motors, Title and Authors

Block Diagram



the above block diagram we are using EEG Electrode. EEG electrode receives the signals from the brain, the received signal is very low so it will be amplified by INA332. It is an instrumental amplifier which amplifies the signal and send that

signal to MSP Board which will verify that signal and decide the purpose of the signal and give to the L293D for driving of motor or any other devices.

1) INA332

It is an instrumental amplifier which amplifies the signal coming from the electrode to readable signal for the MSP Board (Microcontroller) as shown in the above diagram.

2) MSP430F5529 USB Experimenter's Board

It consist of microcontroller, display, micro SD Card Slot, USB Development Platform, etc. MSP Board will receive that signals from INA332 then it will utilize this signal to decide the purpose of signal.

3) L293D

These devices are designed to drive a wide array of inductive loads such as relays, solenoids, DC and bipolar stepping motors, as well as other high-current and high-voltage loads.

4) SERVO

Servo will directly operate through MSP Board

Innovativeness of the Proposed Solution:-

This will be very use full for many fields like in defense, medical uses, industry, also in daily life. This project will increase the human limits. It also make life easy.

Using this concept in defense to control guns, missiles, tanks in critical condition such as war, terrorist attacks etc.

We can use this product as gun controller from long range for the protection of soldier in the battle field & we can use as targeting system.

Now a day we know about the battlefield situation & the terrorist attack it is become very horrible & dangerous because of the latest weapons & advance technology.

But our product will change the scenario of the battle field.

Need of this Project

1) Used for Defense:

Using this concept in defense to control guns, missiles, tanks in critical condition such as war, terrorist attacks etc. We can use this product as gun controller from long range for the protection of soldier in the battle field & we can use as targeting system.

Now a day we know about the battlefield situation & the terrorist attack it is become very horrible & dangerous because of the latest weapons & advance technology.

But our product will change the scenario of the battle field.

1) Use for Patients:

Patients like: - Paralyzed, Coma patients, Handicaps, Prosthetic limbs, etc.

This product can use for the all type of patients those who cannot talk, hear, and walk, blind.

If the patients is in coma the brain of that person is always working. So through this product will help to us to understand what the patient want.

For the paralyzed patient those who can't walk this product can help to drive the wheel chair for those patients using their mind.

This product can use normal person also in their daily life. For make their life easier.

We will make this product low power consumption as well as low cost. Which make this product purchasable for middle class as well as upper class customers.

2. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

3. REFERENCES

1. Taken From the history that is given in the page no. 1 & 2 as well as from IEEE papers the relationship of alpha waves and theta waves in EEG during relaxation and IQ test
<http://ieeexplore.ieee.org/document/5940766/>
2. History:-www.wikipedia.com
3. Figure1:<http://www.braincoreofthelowcountry.com/wp-content/uploads/2014/07/brain-4.jpg>
4. Figure2:- www.neuro24.de/eeg.htm
5. Figure3:- www.neuro24.de/eeg.htm
6. Figure 4, 5, and 6:- <http://www.ti.com>
7. Page 2:- www.wikipedia.com
8. Page 3:- <http://www.ti.com>

Implementing product focused Process improvement through The Post Iteration and Process Improvement Workshop

Tatek Engdashet
Addis Ababa University
IT PhD program
Addis Ababa, Ethiopia

Dida Midekso
Addis Ababa University
Department of computer
science
Addis Ababa, Ethiopia

Ricardo J. Machado
University of Minho,
Department of Information
Systems
Guimarães, Portugal

Abstract: SPI plays a significant role in improving software development performance in a software developing companies. Developing SPI approaches which are contextually easier to integrate with different development environments has been an issue for the research in the area. The SPI framework developed define, the detailed implementation procedures of SPI activities in line with the product development process through the post iteration and process improvement workshop (PIPIW) process steps and SPI tracking model. The framework consists of integrated capability improvement process flow which is developed through the integration of CMMI continuous representation and process components of institutionalization.

The developed process steps developed in the PIPIW is evaluated in a case study sited in a software development setting to examine its applicability for SPI implementation at small development settings. Results of the case study prevailed that, the framework is implementable in line with the development activities. In addition the advantages implementing the framework proven to organize and continually improve the practices used for the development activity. The case study in general prevailed practical evaluation of the framework to demonstrate how it address the perceived requirements.

Keywords: Post iteration and process improvement, SPI, software process, agile methods

1. INTRODUCTION

It is generally accepted that SPI is the crucial factor for the quality of the software product developed in software companies. Continuous improvement of the software process is considered as a means to guide software developing companies towards a standard practices to have a software process with better capability. Standard capability models are created, as a result significant improvement on the quality of the software products, through the introduction of new techniques and new tools, is achieved [1]. Software process capability describes the range of expected results that can be achieved by following a software process. The software process capability of an organization provides one means of predicting the most likely outcomes to be expected from the next software project the organization undertakes [2].

The implementation of SPI programs differs according to the priority of parameters considered for a particular context. Studying the initiatives and their success stories is a valuable input to study and evaluate alternative approach in developing cost effective, simple and context aware SPI framework. Different SPI implementation strategies with their underlined approaches presented in SPI researches. Their focus mainly was the context those companies exist. The software process improvement framework developed at national level is one of the many. The frameworks developed have different features, but basically followed similar procedure. The software improvement framework developed to be followed by Brazilian companies called MPS model [3] is one example. In this framework, implementing software process improvement initiatives in small and medium companies present SPI framework with extended stages than it was in CMMI [4]. The MPS Model is composed of three components. These are the MPS Reference Model; the MPS Assessment Method; and the

MPS Business Model, to guide the process improvement. Mo prosoft [5] is another approach developed in Mexico basically for small and medium software enterprises in Mexico, and later recommended to be used in Latin American software development companies as a national standard. According to the approach, a company should define its process based on the structure of the software organization. It is further recommend that, an organization can attain better success if it define and set up its own process implementation.

Developing SPI model for small and medium companies is the other dimension researched by many scholars and practitioners. One of them is PRISMS [6] which is also an approach to process improvement based on CMMI. It follows an approach to adapt CMMI based SPI implementation, using goal question metrics (GQM), to align process improvement with business values. The other SPI framework is OWPL [7], which is a gradual approach for software process improvement in SMEs. It is based on SPICE and organized with a three stage improvement framework. On the first stage (Stage 1) is called Micro-assessment, where micro-evaluation is done through simplified assessment questionnaire. In the second stage (Stage 2: OWPL evaluation) detail evaluation analysis is achieved which covers 10 process areas followed by the final step (stage 3), which is SPICE assessment. Another approach for SPI implementation in small companies is "Improvement Framework Utilizing Light Weight Assessment and Improvement Planning (iFLAP)" [8]. The approach is focused on selection of any particular KPA(s) and implement the process improvement through selection of professionals for SPI implementation. It start with assessment of the status of capability, then plan improvement program.

In general two important similarities can be extracted from the approaches taken. The first one is selecting and following well

known SPI framework to organize the SPI implementation activities. The other important similarity and recommendation on SPI implementation is, adapting the process to the context of software developing organization. In line with this, SPI activities in small and medium companies recommended to follow development activities. On the other hand, the recently introduced agile methodologies brought a product focused iterative development approach. This opportunity brought a possible path for success in software development in such companies, though long term process improvement is yet to be answered in these methods.

Agile methods have demonstrated a different focus on the direction of software development to be more product focused. However, they have some drawbacks and limitations on their applicability related to nature of project and development environment. The focus of agile methods is on the project only and no formal strategy for long term process improvement is presented [9], [10]. In addition, development activity being mostly dependent on tacit knowledge owned by developers could be a risk if skilled developers leave the organization [11]. Agile practices like XP are sensitive to full understanding and participation of development process. Little documentation is also considered as one limitation of agile practices to trace the development process for developers joining the team and to retain the experience which can be an asset for the organization's future growth [11].

The overall objective of Agile and CMMI approach is ensuring development of a good quality software but from different perspective. The agile methods focus on the "how" aspect of implementing software capitalizing on the people and product development. The CMMI framework in general define requirements a matured software process should have. Focusing on these two important, but considered incompatible characteristic is considered a better approach in recent SPI researches. Such approach can be beneficial for software companies with low maturity level to benefit from their process improvement at the same time keep the pace on the business by responding to customers demand.

In this paper the process improvement guidelines defined by the Post Iteration and Process Improvement Workshop process step is presented in detail. The guideline is part of the component of an integrated SPI framework of agile and CMMI developed earlier [12]. The paper is organized

2. RELATED WORKS

Experience in combining the two practices has shown that CMMI and scrum complement each other by creating synergies. Scrum defines how products are developed which can fill the gap on CMMI where definition of what development method to be used is missing [13]. CMMI, on the other hand, provides guideline and practices to deploy, sustain and continuously improve scrum in organizations [14]. In another study, the combination of CMMI level-5 processes and scrum have also been found complementary [15]. The CMMI framework can specify a clear vision of what processes should be considered for performance improvement and the scrum ensures efficient and flexible implementation of processes. Implementing agile methods can reduce the cost of training and no documentation demand will exist at early stages of software development. Hence, companies can give more focus on software development without losing sight on process improvement [16].

Combining the two approaches has taken the attention of researchers recently. Accordingly, research findings related with combining some agile methodologies with the KPAs in

the CMMI framework have demonstrated the two approaches can work together and can also be even better if they are implemented thoughtfully than they are individually [9, 11, 12, 19]. In line with this, the attempts made and the results obtained have shown tangible evidence of achievements in software development and SPI through combining CMMI and agile methods.

Experience reports from Wake [54] shown that, agile practices are accelerators of SPI with a benefit of providing quality product with time and improve the capability of the organization. Performance improvement has been achieved in both smaller and large projects as a result of using scrum with CMMI. The combination of Agile practices and CMMI is explained as a means to "Amplify Learning and Deliver Fast"[54]. The following summary of related works demonstrate the approaches considered, methodologies used, and findings reported by the researches.

- Application of XP and CMMI v1.2 for SPI implementation for small development firms is presented by M. Yousef et al [17]. According to the research, it is possible to cover twenty (twelve largely and eight partial) out of twenty two KPAs through XP. Unlike other researches [34, 97], the researchers argued coverage of most of the KPAs through XP, though no objective evidence is presented. The description of relationship between CMMI and XP only present at the KPA level and detailed analysis of the specific practices is recommended for future research. The model developed for process improvement is focused on development approaches to follow XP method from agile approach. According to the proposed framework, additional features can be added to XP to fulfill requirement of unaddressed CMMI KPAs. The features to be added can be from the traditional methods according to the research. The process improvement guideline is not presented, rather the XP method is selected to fulfill the development approach.
- Introducing CMMI following the agile methods improve stability of the organization's process while keeping the desired agility in product development [18]. Three companies with different development approach and implemented CMMI and certified to some level. The approach first examine the existing development approach and accordingly tailor the development through introduction of scrum practices. The detailed guideline and common approach for implementation of CMMI and agile method is not the focus. Rather the approach focused on defining a framework where different development entities can introduce scrum to their development culture.
- K. Miler and J.Lukasiewicz [19], developed a reference model called a "CMMI-Scrum (C-S)" model. The model used for mapping specific goals of the second and the third level of maturity in the CMMI staged representation of the CMMI (V 1.2) model onto the activities described by the Scrum methodology. The study consider Scrum to cover some practices of level 2 and 3 of staged representation of CMMI and claim 40% coverage of specific practices. The objective of the approach is to manage the compromise between scrum agility

and CMMI maturity through selection of practices and introducing new ones. The research did not demonstrate the process whereby specific practices or maturity levels improve their capability. The approach is limited to application of scrum in a CMMI environment and identify problems associated in a project than an organized capability improvement plan.

- In another research, M. Pikkarainen, and A. Mäntyniemi [20] presented application of CMMI framework for assessment of agile software development. The research defined relationship between agile practices and CMMI specific goals as a main component. Its main focus is, on assessment aspect, and no clear set of activities with defined target for process improvement is presented. The approach define a model and associated description of the guidelines for process assessment. The requirement to be attained by each agile practice in relation with process capability is not included. The approach of application of CMMI for assessing companies implementing agile practices is discussed, while aiming introducing agile practices to the existing practice, based on project related experience results.
- Application of scrum practices to the CMMI practices implemented environment to introduce agility to a plan driven environment is an approach demonstrated by Ana Sofia, C. Marçal, et al.[21]. The approach taken was focused on aligning practices from scrum to CMMI project management process areas with assessment of their relationships and differences. The main contribution of this research is on the project management process areas on CMMI and scrum practices. The process areas other than project management were not considered. The research concluded that, an improvement on the relationship level between CMMI project management process areas and scrum can be attained through tailoring scrum practices.
- M. Fritzsche and P. Keil [22], presented analysis of all CMMI capability levels with practices from XP and scrum. The detailed analysis presented the compatibility, collaboration and conflict between KPAs from CMMI level 2 to level 5 and practices from XP and scrum. The research concluded that the agile method (XP and scrum) support CMMI level 2 and 3 KPAs. To fully utilize the combination of the relationship between these two sets of practices, a practice catalogue of agile practices is proposed to be studied and developed. The research further recommended analysis of other agile methods regarding their relationship with CMMI. The paper didn't define a guideline or approach recommended to be followed in combining the two approaches or how to implement SPI through combination of practices from the two approaches.

Since the idea is introduced recently, and is coming out of the idea that “agile and CMMI are completely incompatible set of practices”, it needs a thorough investigation in all dimensions to enrich the results achieved so far. The approaches taken mostly are fragmented approach by picking practices from both approaches and considering their complementation to benefit from.

One of the areas which worth considering is aligning the process improvement activities with the development methods of agile. Agile methods execute development through iterations to be completed within 1 to 4 weeks. Extending the iteration workshop to include process improvement activities is considered the possible approach to guide agile methodologies fulfill CMMI requirement of process capability. Based on the activities of the iteration review meetings, process improvement steps can be included to review the experiences and document them through agile based practices templates and procedures. This approach is used to integrate the two paradigms as a unified framework to be used as an alternative path for companies starting product focused SPI implementation.

3. THE POST ITERATION AND PROCESS IMPROVEMENT WORKSHOP (PIPIW)

Please In agile development, at the end of every iteration, a review session is implemented to evaluate the team's performance [23]. According to Cockburn's reflection work shop [23] and PIW by Salo [24].et al. provide project teams a way to examine and accordingly shape the practices while running projects. Moreover, it makes easy to get a quick feedback on the improved practices. The steps used in the post-iteration workshop by Salo et al. is and reflection work shop from Cock burn[] is used as a baseline and contextually extended to incorporate process improvement practices defined in the previous section to define post iteration and process improvement workshop (PIPIW). It starts with the development team discussing the positive and negative experiences of the previous iteration. Based on the identified positive experiences and improvement directions the team identify the practices for the next iteration.

In this research the PIPIW include process improvement steps and guidelines from the integrated capability improvement framework and SPI tracking model defined and explained earlier in our earlier phase of research [12]. The post-iteration workshop technique is used for this research to review activities of the previous iteration of the project and provide a mechanism for the team members to identify positive and negative aspects or difficulties to be improved in the next iteration. The result of practices selected for the next iteration is considered as improvement suggestions in each iteration which in turn can be measured and tracked through the frameworks and associated practices included. The sequence of activities is shown in Figure 1.0 and the detailed description of each indicating their role in the PIPIW.

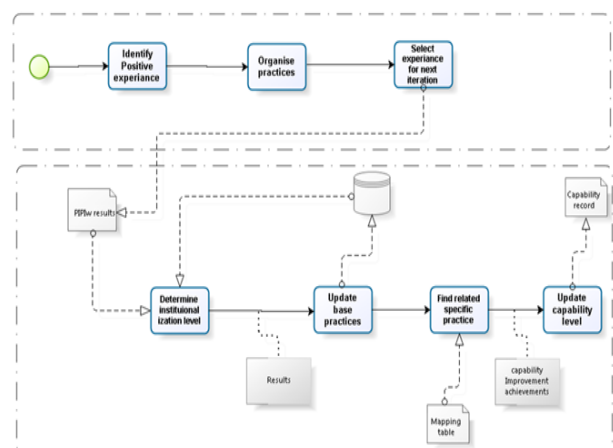


Figure 1.0 Post iteration and process improvement workshop process steps.

Identify and organize positive experiences:

In this step, the development team task is to identify the practices with positive outcomes. The fundamental activities are based on the practices of agile methods (XP and Scrum). The project participants determine the procedures used and templates generated in that specific iteration and explanation of the approach on how positively affect the iteration. This followed by organization of the positive practices according to the nature of the practice or process area where the practice belong. The practice related procedures and templates need to be organized with their associated description. Group discussion is held within the group members reviewing and creating common consensus on the practices identified and organized. In line with this, the group discussion can also review the negative experiences, then the next iteration planning can avoid those practices from happening.

Define activities for next iteration: Based on the collected positive experiences identified in the iteration undertaken by the team and the discussion made, the group members define the activities for the upcoming iteration. The list of practices are recorded on a story board for each iteration to enable comparisons to be made in each iteration to trace the improvement achieved. The different teams can see the practices selected and refined in the iterations carried out by the team.

Similarity Analysis: In this step the agile practices used in each iteration including the updates made is assessed its variation with the previously defined templates and procedures. The previous the process improvement group collect the predefined set of activities defined by the development groups and evaluates their level of similarity. The base practices compiled in the company is used as a bench mark for the analysis and will get updates at the end of each iteration. The resulting output is used to determine the status of institutionalization of those practices in that particular setting. Templates and procedures of each agile practice is measured how they are institutionalized in relation with their respective practices. The level of institutionalization of the practices is defined according to the requirements specified at each level of institutionalization process steps.

Mapping table reference: In this step the task is representing the process improvement in terms of specific practices of CMMI KPAs. Identification of the specific practices is based on the mapping relationship of specific practices of CMMI and agile practices. This is used to identify specific practice of CMMI addressed by the agile practices identified in the previous steps. In doing so, an agile practice can be mapped to different specific practices and a single specific practice can be addressed by more than one agile practice. In cases where more than one agile practice represent a specific practice in a KPA, all the agile practices expected to meet the institutionalization level requirement to improve the capability level of the corresponding specific practice up to a level to be indicated as capability level

Update SPI tracking model: Basically, SPI tracking model is used to follow up and display the improvement progress of KPAs through its specific practices. In the previous steps, the agile practices are evaluated based on their level of institutionalization and identified with which specific practices they can be mapped referring to the mapping table. In the SPI tracking model, the capability level of specific

practices is updated based on their level of institutionalization. The complete capability level of the KPA is represented through the specific practice. This can show where the gap for improvement is; and to consider them in the upcoming capacitation plan of the organization.

The framework for small and medium companies preferred to be more product focused than aiming on organizational process. According to [25], software development business can have more success through continuous improvement, and the packaging of experience for reuse. Hence, the guidelines and architecture of the SPI framework shall be articulated based on this notion. As described by [26], product focused approach to software process improvement emphasizes on defining tools and guidelines for product quality driven process improvement.

The framework basically planned to address the issue of SPI implementation difficulties. In software engineering a number of SPI models and implementation guidelines developed and used for decades. In line with this the implementation success stories and difficulties reported. In addition success factors for SPI implementation prevail considerations to be made in SPI implementation. Understanding the models and their fundamental principle is important to select and define the components.

4. RESEARCH DESIGN

In this section, the research method selected and used, including the research setting is discussed. Based on the target of this research data analysis and evaluation procedures discussed according to the research context.

4.1 Research Method

Case study method is selected to be used for the study based on reasons related to the research context. According to Runeson [27] software engineering researches recommended to follow case study methods to research contexts where the study subject cannot be studied in isolation. In software process improvement when studying a change as a result of introducing an approach case study is the case study is the suitable approach to follow [27]. As pointed out by Runeson [28] “case study research lends itself naturally to software process improvement (SPI) because of the focus of case studies on individual sites within their natural context”. It is also explained by Yin [29] that, case study is a suitable approach to investigate a contemporary phenomenon in a real-life context. Besides, it is a good approach to answer questions related with how and why which are related with operational links to be traced over time [29]. The components with associated guideline of the framework will be used at different levels to collect case study data and guide the analysis procedure. The case study will be executed in a software development environment where the researchers’ role was as participant observer and in some occasions lead meetings in a group discussion.

The case study is targeted to evaluate the PIPIW process steps applicability for SPI implementation, within software development projects. The case for the research is selected according to the definition of case in case study research “contemporary phenomenon in its real-life context” as it is explained by yin [29]. Hence, a software development project can be considered as a study case. The main purpose of the case study is to evaluate the integrated SPI framework in terms of its applicability to improve capability of CMMI KPAs. In this regard, software development environment is the context where the SPI implementation is to be executed.

The study follows case study steps iteratively based on the guides in case study research described by Benbasat as preparing data collection, collecting evidence, analyzing case study evidence, and presenting case studies [30].

4.2 Data Collection

According to Benbasat[30], good quality case study research considers three basic principles, namely use of multiple sources of evidence, creating a case study database, and maintaining chain of evidence. It is generally recommended to use six data sources in conducting case study research which are documentation, archival records, interviews, direct observations, participant observations, and physical artifacts [30]. This research used structured and semi structured individual and group interviews from case project teams' software development documentations, direct observations, and seminars according to the scenario to be investigated.

According to Yin [31] focused group discussion, software development document analysis, individual practitioners interview can be used to keep the data triangulation requirement. According to [27] the use of multiple sources of data in case study research provides a better perspective of the study and increases the credibility of the research. As recommended by Runeson [28] for case study research in software engineering, triangulated data collection methods used in this research. These are documents from project related meeting minutes, records of sprint reviews, and technical documents and the resulting reflection on each of them used. For the interview, discussions is organized with in small groups of software developers and individual developers. For the observation part, process of review meetings, where meeting attendants interact with each other, and thus generate information about the studied object, is used.

4.3 Research Setting

The company where the case study is situated is a software development unit placed under academic institution. The development unit is a separate section founded specifically for application development to internal and external customers. The application development unit has autonomous business orientation in terms of engaging with any software development projects. The financial related tasks are under the hosting institution and clear agreement is established to manage the income associated with software development projects. At the time of the case study the software development department has 10 developers with variety of job positions (a team leader, programmers, system analyst, software architect, security expert). In terms of roles in a specific project all experts take different roles according to the nature of the project to be developed.

The development team started implementing the project following practices from scrum and XP of agile methods with available standard templates used. The source of most of the practices related templates and procedures is from "agilealliance" package. In line with this, the team started organizing the base practices based on scientific definition and procedures of standard methodologies. The base practices are not a complete copy of the standard methodologies but they are selected practices to be used at the beginning of the first iteration. Team members privileged to suggest a template or procedure, for any practice from these methods, based on the practices from the standard methodologies. The improved practices are planned to be updated at the end of each iteration based on the achieved results of the PIPIW.

Based on the procedures of PIPIW, the development team aligned the development task and implementation of SPI activities using the framework. At the end of each iteration the base practices database is updated with any improved template or procedure. The development activity is the primary focus of the team, and it is directed through the daily meeting of sprints. This has contributed to keep focus on the product as a primary job for the practitioners. According to the SPI framework, process improvement is executed at the end of each iteration. Each PIPIW propose practice related procedures or templates for the next iteration or improvement suggestions on the existing ones. The recommendation is based on the experience based practices and enhancements found positively affect the development activities in the completed iteration.

5. DATA ANALYSIS AND FRAMEWORK EVALUATION

Data collection and data analysis in case study research follow an iterative process. In this regard, development of theoretical propositions to guide data collection and analysis is recommended [31]. The analysis and evaluation of the framework is based on the guideline defined in each components of the integrated framework developed. In this research, the data collection and analysis basically focused on the collected results of the development process and process improvement activities with associated templates and procedures. In addition, personal reflections and group discussions have been held based on the interview plan prepared. The interviews and group discussions used to prevail the practitioners view on the application of the SPI framework in line with the development activity. The data collected through interview and group discussions is analyzed following the procedures of qualitative data analysis.

5.1 Discussion of Study Results

The case study results presented are from the compiled result of the case study in application of the integrated framework of agile and CMMI for process improvement. In this part of the thesis, the application of the framework and the results of process improvement is presented after the first project is completed in six iterations. The reported case study result is focused on what has been done in each phases of the PIPIW and the results discovered. In addition, the improvement achieved in due course of implementation of the framework in parallel with the software product development is recorded throughout the case study. The compiled results of the case study is presented in the following section. The output of each iteration is presented by compiling the iterations with similar characteristic than discussing results from each iteration one by one.

First Iteration: - The team started the development task with practices and available templates and general procedures of the standard methodologies considered relevant at the beginning. In the first PIPIW, which is held at the end of the first iteration, some templates and procedures have been found to be relevant to be considered as good experience. The team collected those practices and proposed them to be followed in the upcoming similar tasks. Some of the procedures and templates have been included as part of the initial base practices. But, most of the proposed procedures and templates of the practices implemented, were not considered as part of the base practices. This is done following the decision from group members due to the fact that the development group just started organizing its experience. The development group was set free to consider them for the second iteration or start

developing new templates or procedures for the practices. Following the first iteration, the team organized the practices with related procedures and templates in the base practice table. The table is prepared to show updated status of team document management file to demonstrate how templates and procedures of agile practices are initially started to be organized and updated in the process. Later in the process, such data is planned to get updated at the end of each PIPIW and organized in the database of the PMS. Resources related with development activities are planned to be accessed from the PMS once the software is completed.

Iteration (2 - 6): - Starting from the second iteration, the team started the development by selecting the practices, associated procedures, and templates from the base practice database. In the process of the development, the group used and accordingly made modifications on the procedures and templates. At the end of each iteration, the development group executed the PIPIW according to the predefined procedure. In each PIPIW, templates and procedures of the practices used and proved helpful in the development activities were proposed. The proposed templates and procedures with their associated practices used in the development activities of the next iteration, and modified to incorporate additional features. Every enhancement was done on the templates and procedures, which initially selected by the team from the base practices. The detailed activities and reflections on the achieved results is presented focusing on the main activities performed. The discussion, regarding each process step of the PIPIW summarized for five iterations, is presented in the subsequent paragraphs.

Identify and organize positive experiences: - In this step of the PIPIW, the team proposed practices for the next iteration. In due course of the development task, agile practices have been used according to their relevance for the task. Among the agile practices used, the team managed to develop procedure to be followed, and template to be used to execute the associated agile practice in the next iteration. Starting from the second iteration, the development team reviewed the procedures and templates used in the development activities of the respective iterations. At the end of this phase of the PIPIW, the selected practices selected from the base practice organized with newly created or improved practices and proposed templates. The files related with each practices is organized in a simple MS-word datasheet table. A table is prepared in the development site to keep updated status of team document management file to demonstrate how templates and procedures of agile practices are initially started to be organized and updated in the process. After completion of the last iteration, part of updated procedures and templates with related agile practices in the base practice table is shown in Tale 6.1 and table 6.2.

Table 6.1 Part of XP base practices after six iterations

XP		
Practice	Procedure	[Template]
User story	Link	Link
Metaphor	Link	Link
Coding standards	Link	Link
Continuous	Link	Link
Simple Design	Link	

Table 6.2 Part of scrum base practices after six iterations

Scrum		
Practice	Procedure	[Template]
Sprint Backlog	Link	Link
Product backlog	Link	Link
Sprint Review	Link	Link
Sprint Planning	Link	Link

Define activities for next iteration: - From the practices and associated procedures and templates used, the team (the researcher is participant observer and the team leader who is the department manager organize and led the PIPIW) selected and proposed those which were found helpful to use them in the next iteration. In the second and third iteration, the team proposed additional templates and improvements on the existing ones. As an output to this phase of the PIPIW, the development team defined practices with their template and procedures for the upcoming iteration. The definition is in terms of the improved and newly proposed procedures and templates to be used. At the end of the fifth iteration, procedures and templates (when applicable) is prepared, compiled and filed for almost all the practices used by the development team. Some of the agile practice in the two methods found to be difficult to address through templates and procedures. This is basically related with the limitation of integrating the requirement of those agile practices to the development culture. In the last two iterations, part of the base practices were updated through modifications on the existing procedures and updates on templates proposed by the development team.

The most important case study data is regarding the SPI activities performed based on the guideline defined in the components of the SPI framework. Accordingly the templates and procedures developed and updated in each iteration has been summarized in the discussion. The number of procedures and templates has been represented in the graph in Figure 6.4

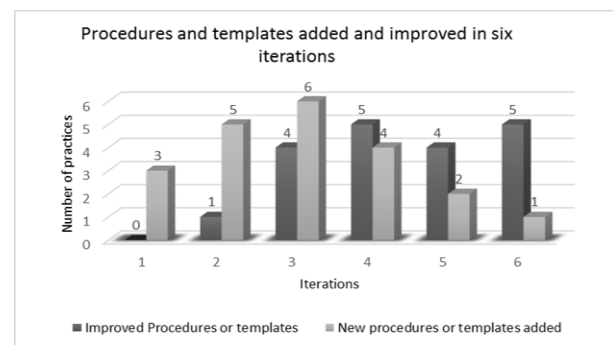


Figure 6.4 Improved and included practices and templates

The graph displays the compiled results of practice related procedures and templates introduced to the base practice database. It also shows improved procedures and templates at the end of each iteration. Templates and procedures were discovered starting from the first iteration and according to the team's PIPIW results some of them are recommended to be considered in the next iterations. The inclusion of new procedures and templates continued till the sixth iteration. The graph also display records regarding improvement on the

procedures and templates on the base practices throughout the six iterations.

Some interesting feature have been prevailed through group discussions and individual interview responses. The developers understanding of SPI implementation has been only outsider look and was considered beyond the reach of small development companies. After implementing the SPI framework the previous perception has been changed. The development team members have developed better understanding and benefits of SPI. In this regard, the SPI implementation being focused at small scale improvement has contributed team members and development unit manager see the benefit of it and encouraged to accept SPI implementation with minimum resistance.

6. CONCLUSION

The implementation of the SPI framework has been executed with minimum, and controlled effort and resource in conjunction with development activities. The implementation of each step in the process of the development activities with the SPI activities is discussed in detail. In each iteration the team has utilized any possible opportunity for improvement on the currently used practice templates and procedures.

7. REFERENCES

- [1] M. Grotke, "Software Process Maturity Model Study," pp. 1-38, 1999.
- [2] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Weber, "Capability Maturity Model for Software , Version 1 . 1," Office, no. February, 1993.
- [3] G. Santos et al., "Implementing Software Process Improvement Initiatives in Small and Medium-Size Enterprises in Brazil," in 6th International Conference on the Quality of Information and Communications Technology (QUATIC 2007), 2007, pp. 187-198.
- [4] A. Regina, M. Montoni, K. C. Weber, E. Edson, and R. D. Araújo, "A Nationwide Program for Software Process Improvement in Brazil," IEEE computer society, pp. 167-176, 2007.
- [5] H. Oktaba, M. Piattini, J. Francisco, and C. Alquicira, "Software Process Improvement The Competisoft Project," IEEE computer society, no. October, pp. 21-28, 2007.
- [6] P. Allen, M. Ramachandran, and H. Abushama, "PRISMS: an Approach to Software Process Improvement for Small to Medium Enterprises," Quality, 2015.
- [7] S. Alexandre, C. Ader, A. Renault, C. R. P. H. Tudor-citi, and F. John, "OWPL: A Gradual Approach for Software Process Improvement In SMEs," Engineering, 2006.
- [8] K. Bos, "Method description : iFLAP," Assessment, no. 3139409, pp. 1-10.
- [9] O. Salo and P. Abrahamsson, "Integrating Agile Software Development and Software Process Improvement: a Longitudinal Case Study," in International Symposium on Empirical Software Engineering, 2005, pp. 193-202.
- [10] D. Turk, R. France, and B. Rumpe, "Limitations of Agile Software Processes," Systems Engineering.
- [11] M. Pikkarainen, O. Salo, and J. Still, "Deploying Agile Practices in Organizations : A Case Study," 2005.
- [12] T. Engdashet, R. J. Machado, and D. Midekso, "Integrated Framework of Agile and CMMI: An Alternative Path towards Product Focused SPI for Small Companies," Lecture Notes in Software Engineering, vol. 4, no. 1, pp. 1-6, 2016.
- [13] Z. Lina and S. Dan, "Research on Combining Scrum with CMMI in Small and Medium Organizations," in International Conference on Computer Science and Electronics Engineering, 2012, pp. 554-557.
- [14] W. Russwurm, "Hidden Treasure : The Implementation of CMMI Practices by Agile Methods," in SEPG, 2010.
- [15] J. Sutherland, D. Ph, and K. Johnson, "Scrum and CMMI Level 5 : The Magic Potion for Code Warriors," Review Literature And Arts Of The Americas, no. November, pp. 1-6, 2005.
- [16] A. Muhammad Ishfaq Khan, Muzaffar Ali Qureshi and Abbas Qaisar, "Agile Methodology in Software Development (SMEs) of Pakistan Software Industry for Successful Software Projects (CMM Framework)," in International Conference on Educational and Network Technology (ICENT 2010), 2010, no. ICENT, pp. 576-580.
- [17] M. Y. Al-tarawneh, M. S. Abdullah, and J. Alostad, "Software Development Process Improvement Framework (SDPIF) for Small Software Development Firms (SSDFs)," International Journal of Computer Science Issues, vol. 10, no. 1, pp. 475-486, 2013.
- [18] S. K. Aggarwal and R. Singh, "Speculation of CMMI in Agile Methodology," in International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 226-230.

- [19] K. Miler J.Lukasiewicz, "Improving agility and discipline of software development with the Scrum and CMMI," in *European Systems and Software Process Improvement and Innovation (EuroSPI)*, 2012, vol. 6, no. February, pp. 416 -422.
- [20] M. Pikkarainen and A. Mäntyniemi, "An Approach for Using CMMI in Agile Software Development Assessments : Experiences from Three Case Studies," in *SPICE*, 2006, no. May.
- [21] A. D. B. Ana Sofia, C. Marçal, Bruno Celso C. de Freitas, Felipe S. Furtado Soares, "Mapping CMMI Project Management Process Areas to SCRUM Practices," in *SEW*, 2007, pp. 13-22.
- [22] M. Fritzsche and P. Keil, "Agile Methods and CMMI : Compatibility or Conflict?," *Software Engineering Journal*, vol. 1, no. 1, pp. 9-26, 2007.
- [23] A. Cockburn, "Agile Software Development : The Cooperative Game," *Building*, pp. 2000 - 2001, 2001.
- [24] O. Salo, "Improving Software Process in Agile Software Development Projects : Results from Two XP Case Studies," in *EUROMICRO'04*, 2004.
- [25] V. R. Basili, G. Caldiera, and H. D. Rombach, "The experience factory," *Science*, vol. 2, pp. 1-19.
- [26] U. Manual, *PROFES User manual*. 1999.
- [27] Per Runeson and Martin Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Software Engineering*, vol. 14, no. 2, pp. 131-164, 2008.
- [28] P. Runeson, M. H. St, A. Rainer, and B. R. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*. United States of America: John Wiley & Sons, Inc, 2012.
- [29] R. K. Yin, *Case study research Design and Methods*, 3rd ed. Delhi India: SAGE Publications, 2003.
- [30] B. I. Benbasat, "The Case Research Strategy in Studies of Information Systems Case Research : Definition," *MIS Quarterly*, vol. 11, no. 3, pp. 369-386, 1987.
- [31] R. K. Yin, *Case Study Research: Design and Methods*, Fourth. SAGE Publications, 2008.