ANALYSIS OF FORCE DYNAMICS FOR AIRFOIL IN THE WAKE OF DIFFERENT GRIDS

Muhammad Ramzan Luhur*, Abdul Latif Manganhar**, Shahid Hussain Siyal***, Khalid Hussain Solangi****

ABSTRACT

This paper presents the lift and drag force dynamics for an airfoil FX 79-W-151A under turbulent wind inflows generated with fractal and classical square grids. The force measurements were performed in a wind tunnel in the wake of these grids. The measured force dynamics is analyzed using classical averaging procedure and a Langevin approach. The comparison of lift and drag dynamics achieved in the wake of fractal and classical square grids, suggests that the fractal square grid contributes significantly extended lift and drag dynamics in terms of standard deviation from the mean as well as short-time local dynamics extracted by Langevin approach. Even at low turbulence level, the fractal square grid wake contributes much higher force fluctuations than the classical square grids.

Key words: *Wind tunnel measurements, force dynamics, airfoil, turbulent inflows, fractal square grid, classical square grid.*

1. INTRODUCTION

Wind turbines operate in dynamic wind fields, which contribute dynamic forces on wind turbine rotor blades. These blades are designed using airfoils that are commonly characterized by static lift and drag forces at fix angles of attack (AOAs) under steady low-turbulence flow [1, 2, 3]. However, in reality, most of the time, wind turbines encounter turbulent inflows due to ground boundary layer effects [4, 5, 6]. The turbulent inflows cause fast variations in AOA [7], which result substantial enlargement in lift and well-known dynamic stall effect [8, 9, 10], leading to large and rapid changing forces on the airfoil.

In order to investigate the dynamic behavior of forces on airfoils, different grids are designed to generate multi-scale turbulent inflows in wind tunnels. These mainly include the classical grids [11, 12, 13] and the fractal grids [14]. Recently, a type of an active grid has also been reported for multi-scale turbulence generation [15]. The classical grids use rectangular arrays of bars and fractal grids use geometric patterns repeated at smaller scales. The type of fractal square grid is experienced to generate more similar wind fluctuations to atmospheric wind than the classical grids [16, 17]. The fractal square grid produces small scale wind statistics similar to those observed in open-air [16].

The aim of this contribution is to provide insight on force dynamics achieved in the wake of fractal square grid and the classical square grids, in particular the comparison of local force fluctuations on short-time scales (high-frequency) obtained with these grids.

The paper is organized as follows. Section 2 describes the measurement details and the analysis approach applied. Section 3 presents the comparison of results from the described grids. The last Section 4 concludes the outcome.

2. METHODOLOGY

2.1. Experimental Data

The force measurements were performed in a wind tunnel for an airfoil FX 79-W-151A at University of Oldenburg Germany [16]. The airfoil was installed in vertical position in closed-loop test section of wind tunnel as shown in Figure 1. The test section has dimensions of 1 m wide, 0.8 m high and 2.6 m long. The mean wind inflow velocity was 50 m/s approximately. The airfoil chord length was 0.2 m leading to Reynolds number of $7x10^5$.

Figure 1: View of the wind tunnel test section. The black arrows indicate the installed position of strain gauge force sensors and the vertical airfoil [16].

* PhD researcher, ForWind-Centre for Wind Energy Research, University of Oldenburg, Germany.

** Assistant Professor, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Pakistan.

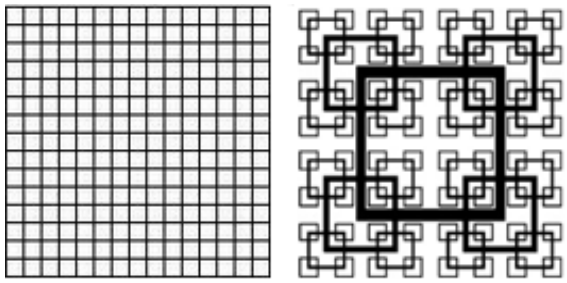
*** PhD Researcher, Royal Institute of Technology (KTH), Stockholm, Sweden.

**** PhD Researcher, University of Malaya, Kuala Lumpur, Malaysia.

The turbulent inflows were generated using fractal and classical square grids as shown in Figure 2. The used grids include; a 5-cm square mesh grid, a 10-cm square mesh grid and a fractal square grid yielding the turbulence intensity of 3.6%, 6.7% and 4.6%, respectively. The turbulence intensity is estimated using the relation.

$$T_i = \frac{\sigma_u}{\overline{u}} \tag{1}$$

Where $\overline{u}$ is the mean wind inflow and σ_u the standard deviation of wind inflow. All three (two different mesh size classical square grids of type as shown in Figure 2(a) and one fractal square grid as shown in Figure 2(b)) grids had a blockage ratio of 25%.



(a) (b)

Figure 2: Types of grids. (a) Classical square mesh grid and (b) fractal square grid.

The forces were measured for AOAs 0°-25° at sampling frequency of 1 kHz using two strain gauge force sensors installed at the end points of airfoil in span-wise direction; see Figure 1. The sensors measure the forces parallel and perpendicular to the wind inflow, which represent the drag and lift forces, respectively. Before measurements, the force sensors were calibrated via balances.

2.2. Analysis Approach

The measured forces are analyzed using classical averaging method and a Langevin approach. The approaches are applied on lift and drag force coefficients obtained using the relations [17]

$$C_L = \frac{F_L}{q.A} \tag{2}$$

$$C_D = \frac{F_D}{q.A} \tag{3}$$

Where F_L and F_D are the measured lift and drag forces respectively, q the dynamic pressure of wind inflow and A the airfoil area.

The classical averaging procedure is applied on C_L and C_D time series at each AOA to estimate the respective mean and standard deviation. Further, to extract more detailed information on force dynamics from this complex turbulence driven system, a Langevin approach [18] is applied in the form of drift function. The drift function in Langevin approach is a deterministic function, also known as first Kramers-Moyal coefficient, which provides full response map of local dynamics of a system. The Langevin approach based on drift function reads [19, 20, 21].

$$D(X,\alpha) = \lim_{\tau\to 0} \frac{\left(X(t+\tau) - X(t)\right)}{\tau} X(t) = X, \alpha \tag{4}$$

Where X represents the C_L or C_D, and α the fix AOA. The approach is applied directly on C_L and C_D times series at each AOA to evaluate the respective mean time derivative of C_L and C_D time series.

Additionally, for better understanding of meaning of drift function and comparison of strength in local dynamics in each grid case, a drift potential is estimated using the relation [17].

$$\phi(X,\alpha) = -\int_0^x D(\overline{X},\alpha)\, d\overline{X} \tag{5}$$

Equation (5) describes the potential in local force dynamics of a system.

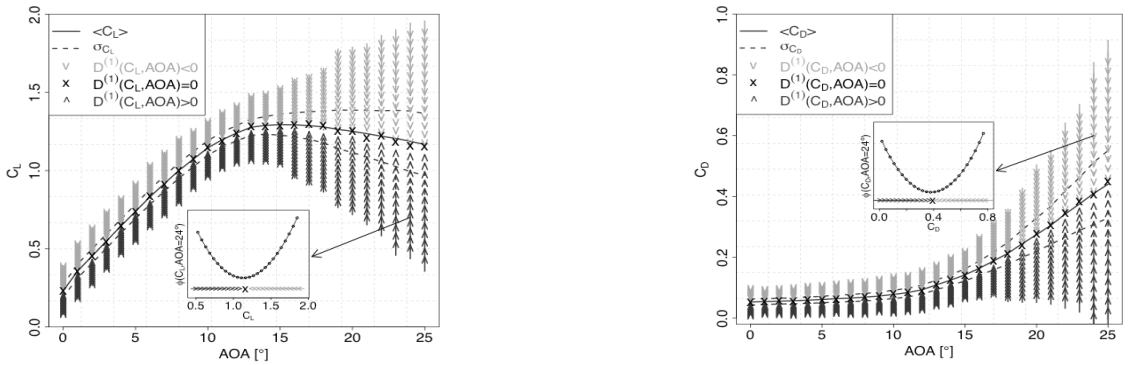
3. RESULTS

The lift and drag force measurements performed in the wake of fractal and classical square grids are analyzed using the approaches described in Section 2.2. The results are presented in terms of static C_L and C_D curves with local dynamics at each AOA. Here, Figure 3 shows the results for fractal square grid, Figure 4 for 5-cm square grid and Figure 5 for 10-cm square grid. In Figures, the mean (solid line) and standard deviation (dashed lines) for C_L and C_D are obtained using classical averaging method, whereas the short-time local dynamics (red and green arrows with black crosses) by Langevin approach using Equation (4). The drift potential is achieved through Equation (5) for AOA 24° in each case. The AOA 24° represents the deep stall regime, which gives better understanding of drift potential with wide local force dynamics compared to drift potential at smaller AOAs.

Comparing the C_L results from Figures 3(a), 4(a) and 5(a), the grid with higher turbulence level (10-cm square grid) shows an increase in maximum C_L with slight shift to higher AOAs. Similar behavior was observed for measurements on NACA 634-421 airfoil by Amandolése and Széchényi [22], and on FX 79-W-151A airfoil by Schneemann et al. [17]. The C_L fluctuations in terms of standard deviation are very small in case of both classical square grids in pre-stall regime and rise gradually as stall

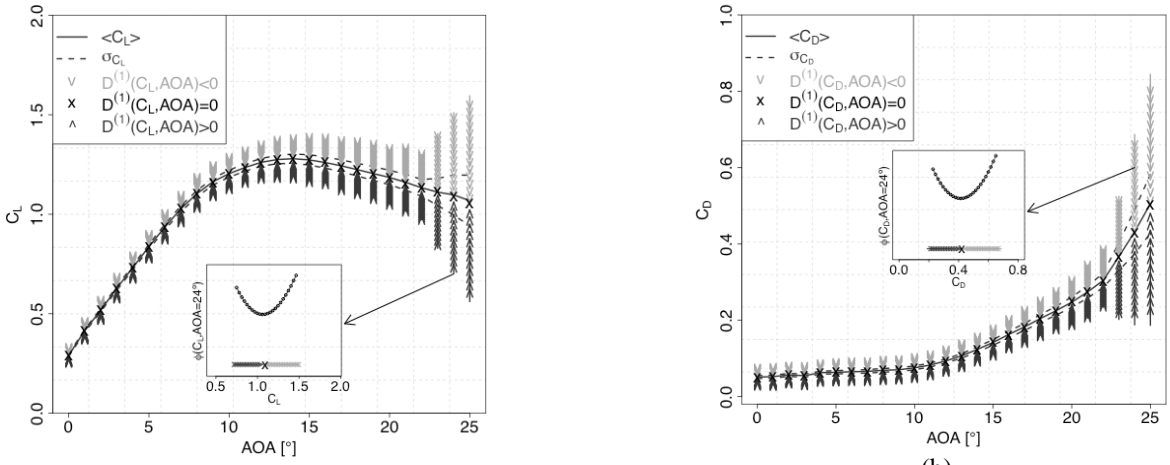
regime starts. In comparison to this, the standard deviations in case of fractal square grid are significantly higher in both regimes despite the lower turbulence level and lower maximum lift than the 10-cm square grid case. The short-time local C_L dynamics based on Langevin approach also show much higher fluctuations with larger potential in case of fractal square grid than both the classical square grids. The Langevin C_L curves consisting of stable fix points (black crosses where drift function is zero) of the local dynamics match almost perfectly with respective mean C_L curves in all three cases.

Again, comparing the case of C_D from Figures 3(b), 4(b) and 5(b), the grid with lower turbulence level (5-cm square grid) shows an increase in maximum C_D at higher AOAs compared to fractal and 10-cm square grid cases. However, the standard deviation in pre-stall regime is slightly lower than both the fractal and 10-cm square grids due to its low turbulence intensity. The C_D standard deviations in the wake of fractal square grid in stall regime are much higher than the classical square grids. Similarly, as in C_L case, the short-time local dynamics of C_D shows remarkable fluctuations with larger potential in case of fractal square grid compared to both classical square grids. The Langevin C_D curves consisting of stable fix points of the local dynamics have almost absolute matching with respective mean C_D curves.



(a) (b)

Figure 3: Static C_L and C_D curves with local dynamics measured in the wake of fractal square grid. (a) C_L versus AOA, where solid blue line represents the mean C_L curve and dashed lines the standard deviation from mean. The red arrows indicate the deterministic increase in C_L and green the deterministic decrease in C_L towards the stable points; the black crosses, where drift function is zero. The black arrow points out the potential in drift function for AOA 24°. (b) C_D versus AOA and same explanation like (a) applies for (b).



(a) (b)

Figure 4: Static C_L and C_D curves with local dynamics measured in the wake of 5-cm square mesh grid. (a) C_L versus AOA, where solid blue line represents the mean C_L curve and dashed lines the standard deviation from mean. The red arrows indicate the deterministic increase in C_L and green the deterministic decrease in C_L towards the stable points; the black crosses, where drift function is zero. The black arrow points out the potential in drift function for AOA 24°. (b) C_D versus AOA and same explanation like (a) applies for (b).

The comparison suggests that the turbulent inflow generated with fractal square grid contributes remarkable extended C_L and C_D dynamics in terms of standard deviation from the mean as well as short-time local dynamics than the classical square grids even at low turbulence intensity.

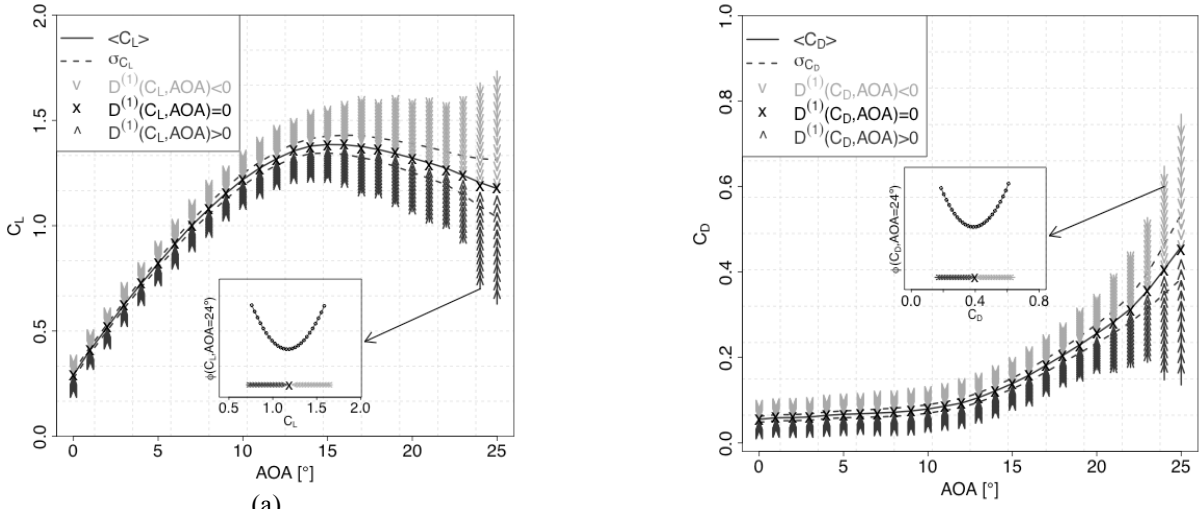


Figure 5: Static C_L and C_D curves with local dynamics measured in the wake of 10-cm square mesh grid. (a) C_L versus AOA, where solid blue line represents the mean C_L curve and dashed lines the standard deviation from mean. The red arrows indicate the deterministic increase in C_L and green the deterministic decrease in C_L towards the stable points; the black crosses, where drift function is zero. The black arrow points out the potential in drift function for AOA 24°. (b) C_D versus AOA and same explanation like (a) applies for (b).

4. CONCLUSIONS

The C_L and C_D dynamics measured in the wake of fractal and classical square grids have been analyzed using classical averaging method and a Langevin approach.

The comparison suggests that the fractal square grid wake contributes remarkable extended C_L and C_D dynamics in terms of standard deviation from the mean and short-time local dynamics extracted by Langevin approach. Even at low turbulence level, the fractal square grid wake contributes much higher force fluctuations than the classical square grids. Further, the variation of AOA has significant effect on C_L and C_D dynamics.

ACKNOWLEDGEMENTS

Authors would like to thank Jörge Schneemann for access to his measured data.

REFERENCES

[1] W. A. Timmer and R. P. J. O. M. van Rooij. Summary of the Delft University wind turbine dedicated airfoils. Journal of Solar Energy Engineering, Vol. 125, No. 4, pp. 488-496, 2003.

[2] P. Fuglsang, K. S. Dahl and I. Antoniou. Wind tunnel tests of the Risø-A1-18, Risø-A1-21 and Risø-A1-24 airfoils. Scientific report Risø-R-1112(EN), Risø National Laboratory, June 1999.

[3] P. Fuglsang, I. Antoniou, K. S. Dahl and H. A. Madsen. Wind tunnel tests of the FFA-W3-241, FFA-W3-301 and NACA 63-430 airfoils. Scientific report Risø-R-1041(EN), Risø National Laboratory, December 1998.

[4] J. Vindel, C. Yague and J. M. Redondo. Structure function analysis and intermittency in the atmospheric boundary layer. Nonlinear Processes in Geophysics, Vol. 15, pp. 915–929, 2008.

[5] N. Troldborg, J. N. Sørensen, R. Mikkelsen and N. N. Sørensen. A simple atmospheric boundary layer model applied to large eddy simulations of wind turbine wakes. Wind Energy, 2013.

[6] J. C. Kaimal and J. J. Finnigan. Atmospheric Boundary Layer Flows–Their Structure and Measurement. Oxford University Press, pp. 287, New York, 1994.

[7] B. Stoevesandt and J. Peinke. Changes in Angle of Attack on Blades in the Turbulent Wind Field. In Proceedings of the EWEC, 2009.

[8] D. M. Eggleston and F. S. Stoddard. Wind turbine engineering design. Van Nostrand Reinhold Company, New York, USA, 1987.

[9] J. G. Leishman. Principles of Helicopter aerodynamics-Cambridge Aerospace Series (2nd edn). Cambridge University Press, Cambridge, 2006.

[10] G. Wolken-Möhlmann, P. Knebel, S. Barth and J. Peinke. Dynamic lift measurements on a FX 79-W-151A airfoil via pressure distribution on the wind tunnel walls. Journal

of Physics: Conference Series, Vol. 75, No. 1, pp. 012026, 2007.

[11] G. K. Batchelor. The Theory of Homogeneous Turbulence. Cambridge University Press, Cambridge, 1953.

[12] S. Corrsin. Turbulence: Experimental methods. Handbook der Physik, Springer, New York, pp. 524, 1963.

[13] G. Comte-Bellot and S. Corrsin. The use of a contraction to improve the isotropy of grid-generated turbulence. Journal of Fluid Mechanics, Vol. 25, pp. 657, 1966.

[14] D. Hurst and J. C. Vassilicos. Scalings and decay of fractal-generated turbulence. Physics of Fluids, Vol. 19, No. 3, pp. 035103, 2007.

[15] S. Weitemeyer, N. Reinke, J. Peinke and M. Hölling. Multi-scale generation of turbulence with fractal grids and an active grid. Fluid Dynamics Research, Vol. 45, pp. 061407, 2013.

[16] M. R. Luhur, J. Peinke, J. Schneemann and M. Wächter. Stochastic modeling of lift and drag dynamics under turbulent wind inflow conditions. Wind Energy, 2014.

[17] J. Schneemann, P. Knebel, P. Milan and J. Peinke. Lift measurements in unsteady flow conditions. Scientific Proceedings of EWEC, Warsaw, Poland, 2010.

[18] H. Risken. The Fokker-Planck Equation-Edition 2. Springer, Berlin, Germany, 1996.

[19] R. Friedrich, J. Peinke and M. R. R. Tabar. Importance of Fluctuations: Complexity in the View of Stochastic Processess. Encyclopedia of Complexity and Systems Science, Springer, Berlin, Germany, 2009.

[20] A. N. Kolmogorov. Über die analytischen methoden in derwahrscheinlichkeitsrechnung. Mathematische Annalen, Vol. 104, No. 1, pp. 415-458, 1931.

[21] J. Gottschall and J. Peinke. On the definition and handling of different drift and diffusion estimates. New Journal of Physics, Vol. 10, No. 8, pp. 083034, 2008.

[22] X. Amandolése and E. Széchényi. Experimental study of the effect of turbulence on a section model blade oscillating in stall. Wind Energy, Vol. 7, pp. 267–282, 2004.

FABRICATION & ANALYSIS OF PORTABLE BATCH-TYPE BIOGAS PLANT

Muhammad Mureed Tunio*, Saleem Raza Samo**, Kishan Chand Mukwana*

ABSTRACT

The growing awareness of pollution problems associated with inadequate management of animal manure & organic waste emphasizes the need for appropriate solution to deal with problem. The biogas plant technology is disseminated for production of methane gas. The study was conducted in order to analyze the production of biogas from organic material by using portable batch-type biogas plant. The biogas process is a complex microbiological process carried out with many steps in anaerobic conditions. In this study cow dung was used as a raw material for anaerobic biogas plant. After degradation of waste the methane gas was recovered at about 18.5 psi with maximum temperature of 40 ^{0}C (digester temperature). The biogas plant capacity was 0.55 m^3, which was found enough to meet the energy requirement of single family (comprising four persons) by using single burner stove. The cost benefit analysis revealed that the cost of gas produced by using biogas plat was just half of the conventional system.

Keywords: *Cow dung , Anaerobic digestion;, Biogas production; Biogas utilisation*

1. INTRODUCTION

Biomass is the only renewable energy source which can deliver electricity, provide heating, cooling and fuel in form of solid, liquid and gas. Biomass supplies more than 11.5% of the world's primary energy and about 79.7% of the world's energy consumption. In 2012, about 194.8 million ton of renewable energy was consumed in the world [1]. Biomass is abundantly available in Pakistan. Around 50,000 tonnes of solid waste, 225,000 tonnes residues of crop and 1.0 million tonnes of animal manure are produced daily. It is estimated that potential production of biogas from livestock residues is 8.8 billion meters^3 of gas per year (equivalent to 55 to 106 TWh of energy). As a result of energy crises, various countries of the world are eager to develop renewable energy. The first Biogas demonstration in Pakistan was installed during the year 1974, with an aim to demonstrate the utility of Biogas technology to policy maker. There after 1976, 21 Chinese type plants were installed on experimental basis in order to assess their working under climatic condition of Pakistan resources [2]. But, the tragedy here in Pakistan, is at present, none of the biogas generation plant is working on commercial scale despite of capital investment of huge amount. Pakistan greatly depends on crude oil as an energy source as during fiscal year 2008-2009, about 8.1 million tonnes was imported. These imports of energy resource have led to incredible import bill of US$ 9.4 billion in the same fiscal year. Thus the Government of Pakistan has laid down various schemes to harness indigenous renewable sources of energy, amongst which, biodiesel, solar and wind and biogas may have a big role to play [2-3].

1.1 Importance of Biogas in Energy Perspective

Biogas is already widely used in developing rural communities. It is created by decomposing waste in an anaerobic environment and piped directly into homes for cooking and heating applications. The anaerobic process converts organic waste into useful energy which is clean, cheap and sustainable. It is fact that biogas is generally environment friendly energy resource that can be used for cooking and heating purposes. The conventional method uses cow dung which is easily accessible to most household inhabitants today [4].

The basic realities of Pakistani society which effect decisions, for instance acceptance of biogas technology, must be given full consideration before making any decision on evolving the technology and designing a campaign for this extension into masses. Pakistan is one of the energy deficit developing countries of the world with all the attendant evils in terms of poor nutrition, weak health, inadequate hygiene, lack of education and insufficient energy availability. These factors demand technologies which make use of locally available raw – materials/manpower and are affordable and also which can be operated by the end – users. Biogas technology fulfils all of these pre-requisites. This technology has gone through sophisticated academic research and now has been reduced to financially affordable and

* Assistant Professor, Energy & Environment Engineering, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah. Email**:** mureed.tunio@gmail.com.

** Professor, Energy & Environment Engineering, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah.

economically feasible level. This has now become, in-fact, one of the most feasible and desirable option for meeting energy needs the rural and semi –urban areas of the country [4-5].

1.2 Importance of Biogas in Environmental Perspective

Two areas are of great environmental concern; first one is the issue of land degradation and deforestation. This concern can be resolved by proper management of sustainable energy crops. However substantial biomass requirement for energy production can be met by using the food industry residues from agriculture and commercial activity along with careful planning of energy cropping. The second issue relates to environmental concerns such as toxic emissions and production of tars and scoots [6].

The countryside population using firewood for cooking utilizes approximately three tones of it annually. This practice not only releases five tones of CO_2 into the air as well as causes air pollution in the households which can cause respiratory infections, inflammation of respiratory tract, lungs disorders, etc. The household use of kerosene liberates about 600 kg of carbon dioxide, whereas with liquid petroleum gas emits 300 kg of carbon dioxide into the air annually. All this can be barred by adopting a cooking system that uses biogas.

Keeping in view the energy crises locally or regionally an anaerobic digester was designed, fabricated and installed in the Energy Park in the premises of Energy and Environment Department of Quaid-E-University of Engineering, Science and Technology Nawabshah.

The Plant was made up of mild steel in which approximately 500 kg waste can be fed for the fermentation process. The capacity is 0.55 m^3. A biological process is taken place for recovery of methane gas commonly known as biogas. It comprises primarily of methane and carbon dioxide that could be used as an alternate fuel for producing electricity at homes and farms particularly in remote areas of Sindh province as well as in Pakistan where electricity is limited. After successful experiment on single burner then a double burner stove was attached to the plant for cooking purpose. Further, the plant reduces methane emissions from the waste, therefore considerable amounts of greenhouse gases can be cut down. The improving of soil quality and the reduction of nitrate erosion protects the ground water by employing this technology.

2. OBJECTIVES

The objectives of this study were focussed on following points.

(i) To generate energy source by using biogas plant

(ii) To utilize fermented slurry as fertilizer for agricultural areas

(iii) Reduction in greenhouse gases

3. SOURCES OF BIOMASS

Biomass is also used to denote products derived from organic matter, that is available on a renewable or recurring basis namely, animal residues (cow dung) , municipal residues, yard waste, agricultural crops and trees, wood and it's residues, plants (including aquatic plants), grasses, food processing waste and manure or excretion. In this regards the cow dung was selected as energy source for production of biogas.

3.1 Biogas

Biogas is produced through the anaerobic biological breakdown of organic matter, such as vegetable matter or cattle dung, at slightly elevated temperatures. , it provides a clean cooking and lighting fuel. It is composed of carbon dioxide and methane and other gases are given in Table 1. Its production process produces slurry that can be used for fertiliser and also kills bacteria, improving community health [1].

Table 1. Chemical Composition of Biogas

Compound	Chemical element	%
Methane	CH_4	50–75
Carbon dioxide	CO_2	25–50
Nitrogen	N_2	0–10
Hydrogen	H_2	0–1
Hydrogen sulfide	H_2S	0–3
Oxygen	O_2	0–2

3.2 Fermentation material Availability

The majority of the material will consist of cattle dung and organic waste produced by Pabal's population. Based upon the gas yields are shown in Table 2 [11].

Table 2. Gas Yields for a Selection of Organic Material

Material	Gas Yield (m^3/kg)
Cattle Dung	0.2
Human Faeces	0.45
Banana Stems	0.75
Eucalyptus Leaves	0.89

In the developing world, cost, technology and resource availability are crucial factors in energy production. Renewable energy sources such as biogas are enabling whole communities to improve their way of life through available, appropriate and cheap energy within rural villages.

3.3 Anaerobic Digestion Technology

Anaerobic digestion is a biological process that converts the organic matter into biogas, biosolids and liquor. Anaerobic digestion is carried out in four processes namely, hydrolysis, acidogenesis, acetogensis and methanogensis [5].

(i) Hydrolysis

Hydrolysis is a chemical reaction during which molecules of water (H_2O) are split into hydrogen cations (H)+ continuously referred as protons and hydroxide anions -(OH) , where large polymers (Plastic) are converted in to simple monomers (Glucose, cellulose and starches).

(ii) Acidogenesis

It is second stage process where simple monomers are converted into volatile fatty acids.

(iii) Acetogensis

It is biological reaction where volatile fatty acids are converted into acetic acid, carbon dioxide, and hydrogen.

(iv) Methanogensis

It is biological reaction where acetates are converted into methane and carbon dioxide, while hydrogen is consumed.

4. METHODOLOGY

Anaerobic digester was designed and manufactured. The material is used mild steel. Initially mild steel as raw material purchased thickness of 10 gauges the size 8" by 4". Then folded and arc welding was carried out in order to fabricate fermentation digester. Then other accessories were mounted necessary parameters were fitted as require.

After manufacturing and installation of portable Batch-type Biogas digester was placed in the premises of Energy and Environment Engineering Department for experimental work and in the first stage substrate (250 kg cow dung and 250 litter water) fed. After three days 5 kg of saw dust as natural catalyst was mixed in order to increase rate of digestion and then at daily basis waste was agitated by agitator manually and it was observed that day by day pressure was gradually increased. After passed twenty days (retention period) approximately pressure was recorded up to 18.7 psi. It has two main parts digester (where the slurry is mixed and fermented to produce the biogas) and other part is gas holder (where the gas is collected and connected to the burner through hose for cooking and lighting purpose. The major components of an aerobic digester are shown in Figure 1.

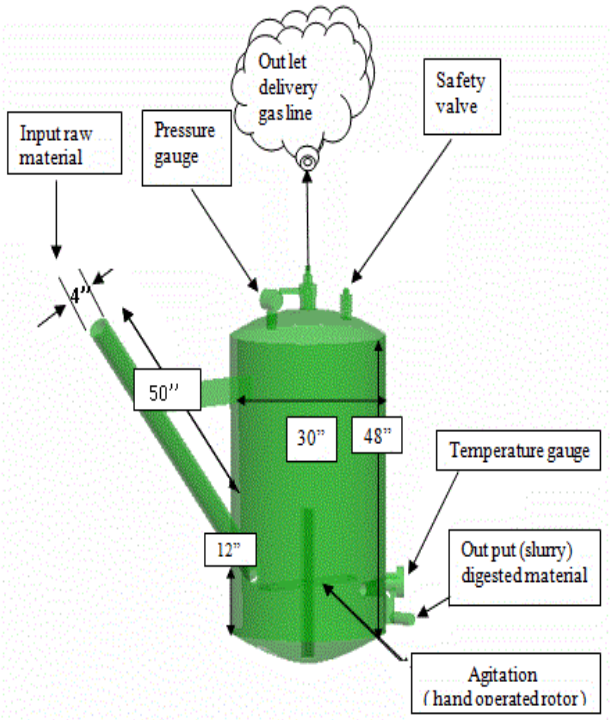


Fig 1: Anaerobic Digester (Batch-Type)

4.1. Operational Parameters of Biogas Plant

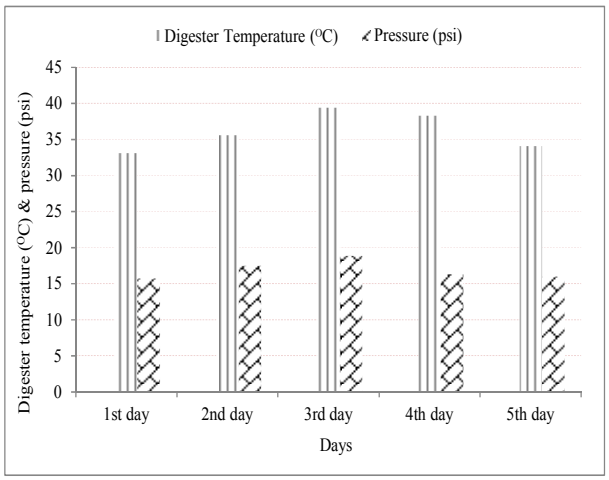
Following parameters play an important role in the production of biogas using different wastes. But it is necessary that these parameters are checked periodically.

(i) Temperature,
(ii) Pressure,
(iii) Solid to moisture ratio,
(iv) pH value,
(v) Feeding rate,
(vi) Carbon to nitrogen (C/N) ratio,
(vii) Seeding of biomass,
(viii) Mixing or Stirring,
(ix) Retention time and
(x) Effect of toxic substances.

5. RESULTS AND DISCUSSION

After completion of retention period tests were carried. It was observed that the due to increase digester temperature proportionally increased digester pressure as shown in Graph 1.The gas yield depends on the ambient temperature and agitation of substrate in the biogas plant. Results also showed that biogas emission consistently increased with three week and diminished during next

five weeks. Emission of biogas became gradually less after fifth week due to declining amount of carbon in the substrate.



Graph 1. Graphical Presentation of Digester Temperature and Pressure V/S Days

When pressure of biogas found maximum 18.5 then a double burner stove was attached to the biogas plant through a pipe for cooking and heating purpose. If biogas is regularly charged with animal waste it continuously will provide biogas supply. The pictorial view is shown in Figure 2.

Fig. 2: Pictorial View of Burning Stove by Biogas

Another result showed that nitrogen (N) content of the compost increased from 1.0 % to fresh of cow dung to 1.8 %. If carbon and nitrogen (C/N) ration is increased of organic martial consequently methane emission will be increased. Some organic materials C/N ratio is shown in Table 4 [10].

Table 4. Carbon and Nitrogen (c/n) Ratio of Various Organic Materials

Raw Materials	C/N Ratio
Cow dung	24
Buffalo dung	23
Horse dung	8
Goat dung	12
Sheep dung	19
Chicken dung	10
Straw (rice)	70
Straw (wheat)	90
Grass cuttings	12
Tree leaves	40

6. APPLICATIONS OF BIOGAS

Biogas has wide ranging applications like electricity generation by using gas generator. It can be used in combine heat and power (CHP) such as gas turbine engine. Additionally it can be used for cooking, space heating. If biogas is compressed it can replaced compressed natural gas for its utilization in automobiles.

7. CONCLUSION

This program was useful for active promotion and dissemination of technologies in remote areas.

It was concluded that portable biogas plant was accessible to shift anywhere as required. It is co concluded that capacity of the plant 0.55 m^3 is quite sufficient for five person's family member. Maximum pressure was produced up to 18.9 psi, which accomplish requirements of single family for 2 hours by using single burner in morning and evening time.

ACKNOWLEDGEMENT

The cooperation and facility provided by the administration of QUEST and laboratory staff of Energy and Environment Engineering Department is highly appreciated.

REFERENCES

[1] Maghanaki, M. M., Ghobadian, B., Najafi, G., and Galogah, R. J., "Potential of biogas production in Iran", Renewable and Sustainable Energy Reviews, 28, 702-714, 2013.

[2] Chakrabarti, M. H., Ali, M., Usmani, J. N., Khan, N. A., Hasan, D. U. B., Islam, M. S., and Irfan, M. F., "Status of biodiesel research and development in Pakistan", Renewable and Sustainable Energy Reviews, 16(7), 4396-4405, 2012.

[3] Dr. Mike Clifford "Appropriate Technology: Biogas, Design and development of a biogas compression and storage system capable of implementation in the developing world" Department of Mechanical, Materials & Manufacturing Engineering the University of Nittingham (2011).

[4] Manga, V. E., Forton, O. T., Mofor, L. A., and Woodard, R., "Health care waste management in Cameroon: A case study from the Southwestern Region", Resources, Conservation and Recycling, 57, 108-116, 2011.

[5] K. Stamatelatou, G. Antonopoulou and G. lyberatos "Production of biogas via anaerobic digestion", Handbook of Biofuels Production, Processes and Technologies, A volume in Wood head Publishing Series in Energy 2011, Pages 266–304

[6] Pöschl, M., Ward, S., and Owende, P., "Evaluation of energy efficiency of various biogas production and utilization pathways", Applied Energy, 87(11), 3305-3321, 2010.

[7] Ducom, G., Radu-Tirnoveanu, D., Pascual, C., Benadda, B., and Germain, P., "Biogas–Municipal solid waste incinerator bottom ash interactions: Sulphur compounds removal", Journal of hazardous materials, 166(2), 1102-1108, 2009.

[8] Alternate Energy Board Islamabad "Biogas Technology by Appropriate Technology Development Organization", 2006.

[9] Amrit B. Karki and Kunda Dixit. "Biogas field Book" By Nepal. 2008.

[10] Kothari, D. P., Singal, K. C., and Ranjan, R., "Renewable energy sources and emerging technologies" Prentices Hall of India (PHI) Learning Pvt. Ltd, 2010.

[11] Tiwari, G. N., and Ghosal, M. K., "Fundamentals of renewable energy sources. Alpha Science International Limited, 2007

[12] B.H. Khan "Non-Conventional Energy Resources" Tatta Mc GrawHill and Company, 2010.

OPTIMIZATION OF PROCESS PARAMETERS FOR TEMPERATURE DISTRIBUTION IN ORTHOGONAL METAL CUTTING

Hassan Farooq*, Mirza Jahanzaib**

ABSTRACT

Process parameters demands optimal conditions to achieve surface finish and desired quality level. The temperature produced during cutting plays significant role in the machining and directly influences process parameters and surface finish of work piece. An investigation has been carried out to see impact of temperature at different process parameters using mild steel AISI 1018 with HSS (Co 10%) cutting tool using a Tool Work thermocouple technique. The analysis is done by using the Taguchi method which optimizes different machining parameters used in turning process for optimizing temperature of tool-chip interface. The parameters are also modeled in GA method and coded in MATLAB. The results obtained from the Taguchi's method shows that the cutting speed comes out to be the most significant parameter among the parameters. The results obtained from Taguchi method are also evaluated by ANOVA.

Keywords: *Process Parameters; Turning, Optimal Seeking; Taguchi Method; GA method*

1. INTRODUCTION

A lot of heat is generated during machining process when the deformation of tool or work piece takes place. The temperature is termed as tool temperature which is produced on the surface of tool when cutting tool comes in contact with the job. Temperature is a parameter that affects the tool performance during the machining operation. High level of heat generation takes place in the result of power consumed during machining.

Rise in the temperature during machining operations is important because temperature depends upon the cutting forces and frictional forces. During machining, approximately most of the heat produced is converted into job material and tool. Thus the power losses remain an important issue to control it. High temperature produced during machining badly affect on tool life and desired quality of product. High temperature increases tool wear and hence surface roughness of job so to reduce cutting temperature during machining is very important objective of this research work.

To control the parameters affecting on machining temperature is significant. Mainly there are major three very distinct zones where heat is produced,(1) shear zone or the primary deformation zone, at this zone plastic deformation occurs,(2) the second zone is tool-chip interface and here the secondary plastic distortion occurs because of frictional rubbing between cutting tool and the heated chip, (3) there is another third zone where heat is produced at work-tool interface, actually this is at the flanks where the strong frictional rubbing occurs.

Flux of heat generated at the secondary zone influences the contacting mechanism by varying situations of the frictional forces, in a result it imparts its role on the position and shape of primary and secondary zone. In the secondary zone, the rise in temperature is a result of power consumed in the deformation of metal chips. The final zone, heat produced at tertiary zone because of work done against forces to overcome the rubbing friction.

Tool work thermocouple K-Type has been used for measuring temperature. A designed approach of Taguchi Method is adopted for seeking optimality and a GA code has been developed in MATLAB to compare the results. The upcoming section presents a brief literature related to previous work followed by experimentation, design of experiments, results and GA optimization.

2. LITERATURE REVIEW

Temperature affects the process parameters during the machining operation. The high temperature in machining influences on surface finish and quality of product. . High level of heat generation takes place in the result of power consumed during machining and transformed into the heat flux. Some experimental methods have been reviewed to find the heat involved in machining operations [1]. Accurate temperature measurement is a challenging task to find the effect of process parameters involved [2]. The rise in the temperature during machining operations is important because temperature depends upon the tool-chip interface [3]. It has been learnt that cutting and frictional forces between the cutting tool and machined chip also play a significant role [4].

* Research Associate, Department of Industrial Engineering UET Taxila, Pakistan

** Associate Professor, Department of Industrial Engineering UET Taxila, Pakistan

Different factors affect on cutting temperature like cutting speed [3], Feed [5], nose radius [4], depth of cut [3], proper lubrication and Tool angles [12]. It was investigated that mechanism of chip formation and dispersion during machining affects on machining temperatures [5], these parameters play important role in the thermal distortion of the machining tool that is supposed to be the major cause of error in the cutting operation [6]. It was learnt that out of all these parameters speed, feed and depth of cut were more influencing factors on cutting temperature as discussed by A. Fata[3], S. M. Cotterell (2013) [9] and N.Abukhashim et. al. [21].

Three distinct zones where heat is produced the shear zone or the primary deformation zone, at this zone plastic deformation occurs [7], the second is the chip-tool interface zone and here the secondary plastic deformation occurs because of the force of friction due to rubbing is produced between the cutting tool and the heated chips takes place [2] and there is another third zone where heat is produced is at the work-tool interface, actually this is at the flanks where the strong frictional rubbing occurs [8].

Also the high temperature of the job material imparts an important role in the primary deformation zone [9] as it softens the material; ultimately the cutting forces as well as the energy which is required to do the further shear [10].

It is seen that the heat generation in primary as well as in the secondary zones strongly depends upon cutting conditions [11] on the other hand heat produced in tertiary deformation zone is highly affected by tool flank wear and tear [12].It is noted that heat generation and the power consumption in metal cutting machining operation depends upon properties of job material [13] and tool material i.e. chemical and the physical properties [14] the tool geometry of cutting tool and the related cutting conditions [15].

Flux of heat generated at the tool–chip interface [16] influences the contact mechanism by varying situations of the frictional forces [17] ,in a result it imparts its role on the position and shape of primary and secondary deformation zones [18]. The rise in temperature in the secondary zone is due to power dissipation in deforming the metal chips [12] which are generated due to the rubbing friction. The final zone, the heat produced at tertiary zone because of work done against forces to overcome the friction [19] occurs [20, 21].

To measure the temperature along the tool, there are different techniques which have been developed [18]. The major techniques which have been developed to measure temperature during cutting are metallographic, Infra red radiation and thermocouple [22].

The embedded thermocouple method is difficult to perform because the holes are difficult to machine, e.g. carbides, High Speed Steel tools and ceramics also required thermocouple is far from point of contact so it is a disadvantage. The hole made by EDM is also at a distant from the contact point that don't allow proper readings as it is limitation because of tool wear and tear the hole damage [18].

Another technique was developed by Sekulić, Gostimirović et al. [22] to measure the temperature along tool using a cutting tool of two materials (sandwich) consists of very fine powder having a definite melting point (constant M.P.). Metallographic is not commonly used method it involves to check and identify the boundary between the melted and un melted. The temperature can easily be found depending upon nature of material and its melting point. R. Komanduri [18] used a thermal radiation micro detector to calculate the temperature in elasto hydrodynamic. In this method the radiation's intensity is detected and measure after bouncing back from the surface of material under consideration.

In this research work K-Type thermocouple is used, the temperature measuring range of K type is up to 1250 ^{o}C and it works on seebak effect where the tool acts as hot junction and work piece as cold junction and an electromotive emf establish in between hot-cold junction as discussed by Sekulić, Gostimirović et al. [22] .

The analysis carried out by Taguchi method for optimization of process parameters. This method is good approach for the robust design which gives rise to the different combination and the set of factors as well as levels in an efficient way that reduces the time and experimental cost [22].

There are many optimization techniques which had been studied and implemented by researchers. Response surface methodology had used for process optimization by R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook(2009) [23] similarly benefits and drawbacks of Factorial design were discussed in detail by M. E. Capella-Peiró et. al (2006) [27]. Unlike Factorial and Response Surface methodology Taguchi approach to design of experiments in easy to adopt and apply for users with limited knowledge of statistics, hence gained wide popularity in the engineering and scientific community. The Taguchi method is a famous optimization technique that provides a very efficient and precise methodology for process optimization and this is a powerful tool for the design of high quality systems. This is an engineering methodology for obtaining product and process condition, which are minimally sensitive to the various causes of variation, and which produce high-quality products with low development and manufacturing costs [22]. This

technique reduce experimentation time and hence product cost. Signal to noise ratio and orthogonal array are two major tools used in robust design.

Taguchi results have been compared with GA method. Genetic algorithm is used to find out optimized function variables and objective function using MATLAB. The Taguchi method is evaluated using the software Minitab 15. The cutting temperature is experimentally measured during turning of AISI 1018 Mild steel and HSS tool by using tool work thermocouple method. The process parameters evaluated are speed, feed and depth of cut. Results of Taguchi method are compared with Genetic Algorithm method. ANOVA is carried out to analyze the effects of machining parameters on the cutting temperature.

3. EXPERIMENTAL DETAIL

The aim is to optimize cutting parameters for low tool chip interface temperature. The experiments were designed on the basis of Taguchi method. The analysis of experiment data is carried out using S/N ratio analysis and ANOVA.

To find optimum machining conditions, signal to noise ratio is used. The ANOVA is performed to find the percentage contribution process parameters on machining temperature.

3.1 Work piece material

For investigation AISI 1018 steel was selected. AISI 1018 was identified due to its broad usage and machining applications in the field of manufacturing industries [4]. The chemical composition of AISI 1018 steel is given in the Table-1. The length and diameter of the work piece was 160mm and 25.5mm respectively as shown in Fig.1

Table-1 Chemical composition of AISI 1018 work piece (in %age)

Carbon	Iron	Manganese	Phosphorus	Sulphur
0.20	99.26	0.90	≤ 0.04 %	≤ 0.05

Fig.1 Work piece material for machining

3.2 Experimental procedure

The experiments were conducted on Lathe with a variable speed between 90 and 150 m/min with a power rating of 12kw. Prior to actual machining, sand paper was used to remove rust layers from the work piece.

The cutting tool used was high speed steel (HSS) tool. Temperature measurements were carried out using tool-work thermocouple method. EDM machine is used to drill a hole of 1mm in HSS tool and a K-type thermocouple of 0.8 mm was seated inside HSS tool as shown in Fig.3. The thermocouple was positioned in 1 mm hole made inside the tool by using Electric discharge machining. The EDM die sinker machine used for carrying out hole is shown in Fig.2

Fig.2 Electrical Discharge Machine (EDM) Die Sinker

Fig.3 Thermocouple arrangements on tool

The experiments were carried out under dry cutting conditions. Dry cutting is selected as use of the oil based cooling lubricant/fluid has proven to be one of the most unsustainable element of the machining process [12]. Furthermore, steel work piece is considered as a good

candidate for dry machining [17]. The schematic of the thermocouple arrangement on tool as shown in the Fig.4

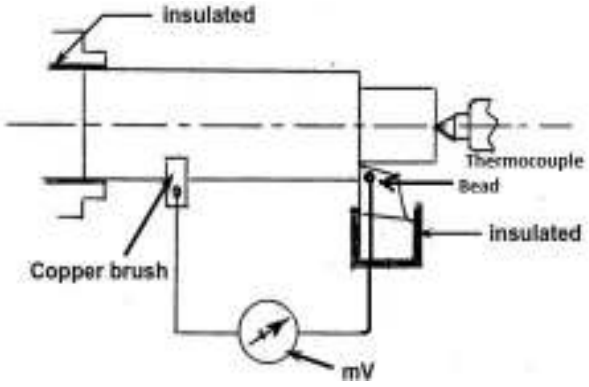


Fig.4 Schematic of experimental Procedure

3.3 Experimental Plan

For turning different machining parameters affect on cutting temperature like cutting speed, nose radius, feed, lubricants, coated and uncoated tools etc. Out of these three main factors feed, speed and DOC have been selected as these factors impart major role in increasing cutting temperature [3]. For turning AISI 1018 steel by using high speed steel, machining parameters ranges were selected as per the standard published by Sandvik Catalogue.

Experiments were planned on the basis of Taguchi orthogonal array that reduces extra effort and the number of experiments. The designed experiments were conducted using Taguchi approach according to a three level, three factors orthogonal array $L_9(3^4)$.

The experiments were conducted at three different cutting speeds (90, 120 and 150m/min), three different feed rate (0.09, 0.12 and 0.15mm/rev) and three different depth of cut (0.30, 0.45 and 0.6mm).The cutting factors and their levels are shown in Table-2. The experimental layout of the L9 orthogonal array is shown in Table-3.

Table-2. Designed parameters at 3 three levels

Coded symbol	Factors	Levels		
		*L1	*L 2	*L 3
A	Cutting speed v, m/min	90	120	150
B	Feed f, mm/rev	.09	.12	.15
C	Depth of Cut d, mm	.3	.45	.6

*L1, *L2, *L3 are 1st ,2nd and 3rd level of cutting parameters

Table-3 Experimental layout using L_9 orthogonal array

Experiments	Turning parameters levels		
	v (m/min)	f (mm/rev)	D (mm)
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

3.4 Analysis of S/N ratio

Signal to Noise ratio denoted by S/N is used to find the response deviating from the desired value. The term Signal represents the desirable value (mean) for response and the term noise represents undesirable value (standard deviation). So, ratio of mean and standard deviation is expressed in term of S/N ratio. S/N ratio can be measured as:

$$\eta = -10log(M.S.D.)$$

Here M.S.D mean standard deviation. The goal of this research is to optimize the turning parameters so that low temperature can be achieved. Thus the observed value of temperature was set to minimum. This means the objective function, S/N ratio is based on smaller the better characteristic as given by the following equation discussed by Yang and Tarng [4].

$$M.S.D. = \frac{1}{m}\sum_{i=1}^{m} T_i^2$$

Here m=no. of the experiments, T_i^2 shows the value of the tool chip interface temperature for the ith experiment. The S/N ratio estimated from experimental data was used to measure and predict response at the optimal level. S/N ratio of response variable at optimum level $\hat{\eta}$ was calculated by the following equation discussed by Yang and Trang[4].

$$\hat{\eta} = \eta_m + \sum_{i=1}^{k}(\eta_i - \eta_m)$$

η_m = mean S/N ratio, η_i = S/N ratio at the optimal level, k = no. of parameters that affect on response.

3.5 ANOVA:

Analysis of variance was carried out significant factors can be identified that influence on response. The following equation is used to calculate total sum of squared deviation.

$$SST = \sum_{i=1}^{n} (\eta_i - \eta_m)^2$$

n= no. of experiments

Taguchi orthogonal array is used to perform the experiments and the experiments were planned and implemented using orthogonal array $L_9(3^4)$. Taking orthogonal arrays and corresponding 3 levels (coded by:1,2,3) as shown in Table 2. In the Taguchi orthogonal arrays the selected factors are leveled and best chosen orthogonal array in this experiment is L_9 containing 3 number of factors and 3 levels. These levels have been carefully selected by considering hardness of material and its influence on cutting parameters.

The principle of thermocouple is based on the (Seebak Effect) which states, an electromotive force (emf) is set in the circuit when two different metals come in contact with each other by conducting media provided that there should be a hot junction as well as a cold one. Thus electromotive force is generated in the circuit due to temperatures difference of both cold and hot junction.

Most commonly used method is tool work thermocouple [22]. Experimental setup and conditions are discussed here. The lathe was used for turning operation and the AISI 1018 MS round bar work piece of 160mm in length. For temperature measurement the tool work K type thermocouple (ranges up to 1250 ^{o}C) is mounted on the work piece.

There is a difficulty of tool work thermocouple to set up an accurate junction between the pair of work piece and tool for establishment of correct readings. From mV after mounting it on thermocouple with the cutting tool, average cutting temperature is evaluated. The thermocouple is calibrated so that a proper relation could be established between tool and work piece. In this case tool acts as hot junction while work piece acts as the cold junction.

Standard thermocouples monitored junction temperature directly [AlCr-Al]. During machining operation, the determination of temperature distribution on tool is included in set of experiments. The set of experiment carried out using Taguchi approach hence combinations of parameters at their different levels are formulated as discussed in Table-4.

Table-4. Design Matrix

No	Factors			Parameters			T , ^{o}C	S/N
	v	f	d	v m/min	f mm/rev	d mm		
1	1	1	1	90	0.09	0.3	298	-49.4843
2	1	2	2	90	0.12	0.45	300.6	-49.5598
3	1	3	3	90	0.15	0.6	298.1	-49.4978
4	2	1	2	120	0.09	0.45	297.8	-49.4785
5	2	2	3	120	0.12	0.6	298.9	-49.5105
6	2	3	1	120	0.15	0.3	300.6	-49.5598
7	3	1	3	150	0.09	0.6	299.3	-49.5021
8	3	2	1	150	0.12	0.3	305.2	-49.6917
9	3	3	2	150	0.15	0.45	301.6	-49.5986

There are 3 levels of each process parameter and mean signal to noise ratio is calculated against each level. Total nine experiments were performed and the S/N ratio for each experiment was also found then listed in graph for machining temperature. Signal to Noise ratio corresponds to lower variance of output characteristics near desired value.

The S/N ratio characteristics can be divided into three categories when the characteristic is continuous: nominal is the best, smaller the better and larger is better characteristics. For the minimal cutting temperature, the solution is “smaller is better”, and S/N ratio is determined according to the following

equation 4:

$$\frac{S}{N} = -10log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) \quad \text{.....Equation: 4}$$

where n is the number of replication and yi is the measured value of output variable. The “minimal temperature” can be achieved using the cutting parameters where S/N ratio is “maximal”. In fig.5 it is clear that on x-axis there are cutting parameters in increasing order from 90-150,0.09-0.15 and 0.3-0.45 similarly on y-axis S/N ratio values in ascending order up

to the minus 49.50. Hence maximal value of S/N ratio in Fig.5 is at 90 m/min(1st level of speed,) at 0.09mm/rev(1st level of feed) and at 0.6mm(3rd level of depth of cut)

The value of cutting speed decreases and up to the 150, the larger value of S/N ratio corresponding to speed is 90 m/min cutting speed. So the optimal spindle speed is at level 1.The feed's influence on machining temperature, is decreasing up to the level 2 then it increase again up to the level 3. So the optimum feed corresponding to S/N value is at level 1 (0.09 mm/rev). Similarly effect DOC value is increasing up to the level 3. The optimum DOC is at level 3 (0.6 mm). The influence of turning factors on the cutting temperature for S/N ratio is given in Fig.5

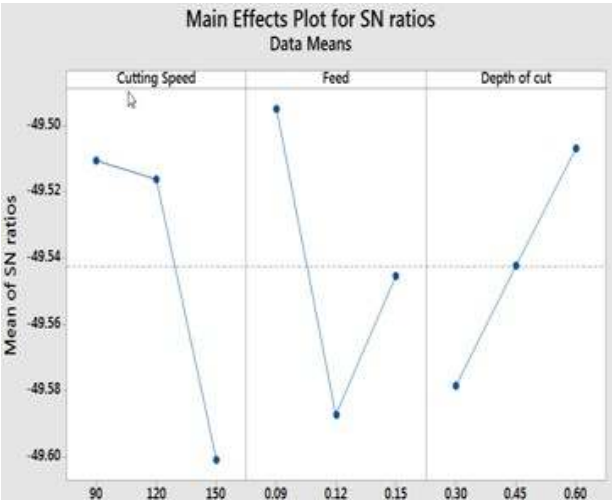


Fig.5 S/N Response Graph for cutting parameters

4. RESULTS AND DISCUSSIONS:

From the S/N response graphs as shown in the Fig.5, optimum process parameters can be found. In graph desired temperature is at the higher Signal to Noise values. The value of speed at its first level (90 m/min), value of feed at its 1st level (0.9 mm/rev) and value of DOC at its 3rd level are selected to be the optimal parameters.

ANOVA test was performed on the data obtained from the experimental results to assure and identify control parameters that influence on the performance measures. After performing ANOVA results are obtained shown in Fig.6. In this ANOVA analysis was done at α=0.10 (significance level) i.e. 90% confidence level. Also those factors have significant contribution to performance measure having P-value< 0.10.

It is evident that the P-value of depth of cut (0.074) and is less than (α=0.10) while the P-value of cutting speed i.e. 0.062 is least. So speed has major effect on performance measure on the other hand feed has least effect. The detailed ANOVA results are mentioned in Table 4.1.

Use your mouse to right click on individual cells for definitions.

Response 1 R1

ANOVA for selected factorial model

Analysis of variance table [Classical sum of squares - Type II]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	0.036	6	5.979E-003	11.66	0.0811
A-CS	0.015	2	7.654E-003	14.93	0.0628
B-DOC	0.013	2	6.417E-003	12.51	0.0740
C-F	7.733E-003	2	3.867E-003	7.54	0.1171
Residual	1.026E-003	2	5.128E-004		
Cor Total	0.037	8			

Fig. 6 Analysis Of Variance of work surface temperature by Software DX 8.

Note: *A-CS = cutting speed, *B-DOC=depth of cut, *C-F=feed , *df=Degree of freedom

Table. 4.1 Analysis of Variance (ANOVA)

Factors	DOF	Sum of Squares	Mean Square	F-value	P-value
speed	2	50.469	25.234	14.93	0.062
DOC	2	42.229	21.114	12.51	0.074
feed	2	26.462	13.231	7.54	0.117
ERROR	2	3.429	1.714		
TOTAL	8	122.6			
S = 1.3 R-sq = 96.90 % R-sq(adj) = 90%					

It is evident that P-value of speed and DOC are less than 0.10 explaining that the speed and DOC has significant effect on temperature. Also the rank or the degree of contribution of the parameters at their different levels is the speed ≥ depth of cut ≥ feed respectively.

When ANOVA test performed using Design Software DX.9 (shown in fig.6) , the linear equation resulted by regression analysis. The relationship of factors by multiple linear regressions is obtained:

Temperature T = 294.020+0.052*v+29*F-8.311*DOC
Eq……..4.1.

Temperature T =294.020+0.052*90+29*.09-8.311*0.6
=298.06

Where v is cutting speed, F is feed and DOC is depth of cut. It is evident from above results that out of all cutting parameters, cutting speed is most contributing factor to the response. Hence Optimal setting of control factors are obtained. From Fig.5 optimal values of control factors obtained from Taguchi method i.e. speed, feed and DOC are put into the multiple linear equation 4.1. Thus the temperature 298.06 obtained which is our desired temperature. Results obtained from ANOVA closely match with Taguchi method.

Thus the optimal setting of control parameters is arranged in Table-5.

Table-5 Optimal setting of control Factors

Control Parameters	Levels	Parameter Setting
Cutting Speed v, m/min	1	90
Feed f, mm/rev	1	0.09
Depth of Cut d, mm	3	0.6

5. GENETIC ALGORITHM METHOD

The GA is used to validate the results obtained from the Taguchi method and the cutting parameters with lower bound to upper bound range of cutting speed 90-150 m/min, range of Depth of cut is 0.3-0.6 mm and Feed range 0.09-0.15 mm/rev are under consideration respectively. The chromosome is defined as the chromosome consists of 3 genes (cutting speed, feed and depth of cut) Our variable bound are:

90 m/min ≤ v ≤ 150 m/min

0.3mm ≤ DOC ≤ 0.6 mm

0.09 mm/rev ≤ f ≤0. 15 mm/rev

$$x_{il} \leq x_i \leq x_{iu}$$

Where x_{il} lower bound and x_{iu} is upper bound

Table-6. Process parameters used in experimentation

Sr. #	Factors	Units	Level 1 (L 1)	Level 2 (L 2)	Level 3 (L 3)
1	Speed, v	m/min	90	120	150
2	Depth of Cut, d	mm	0.3	0.45	0.6
3	Feed, f	mm/rev	0.09	0.12	0.15

Genetic algorithm is a good method to find out the individual solutions out of random population. Population size varies and the size variation depends upon the nature of the problem, also it is interesting that it give rise to hundreds or thousands of possible solutions. GA method coded in MATLAB. The Fig.7 presents generation sequence and fitness value. There are Replication of Chromosomal generations (random experimentations) on abscissa while on ordinate fitness values (optimized values) are mentioned. It is found that the best fitness value (optimized temperature) after several experimentation obtained is 298.5 which is closer to that obtained from Taguchi method i.e. 297.8 It comes out that from the generation(up to 51 experiments), 298.5°C is best fitness value (optimal value, denoted by blue dots) with a mean of 296.334.

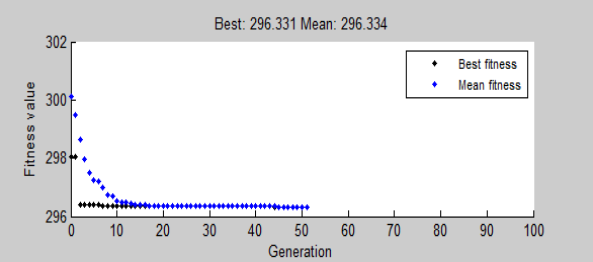


Fig.7 Fitness value versus Generation plot

From the GA method the final points corresponding to the different levels of cutting parameters are identified as 90, 0.09 and 0.6 respectively and our objective function is 298.5.The objective function is basically multiple linear regression equation 4.1 at the optimized levels of function variables i.e. first level of cutting speed 90 m/min, first level of feed .09 mm/rev and third level of depth of cut 0.6 mm which is met (as shown in fig.8).

The Fitness value vs Generation values plot of Fig.7 shows that there are the different fitness values and the curve consists of blue dots showing the mean fitness values while on the other hand the line containing the small black dots representing the best fitness values.

In the Fig.8 on x-axis there is individual's population (no. of factors or variables) while on y-axis best individuals are mentioned. In this plot our individual population/ variables on x-axis are (1) speed (2) Depth of cut and (3) feed respectively. And on y-axis first column goes up to 90, 2nd smaller column up to 0.3and 3rd smallest goes up to 0.12. These three values of parameters are closest to the values obtained from Taguchi method (fig.5). Hence it is validated that the optimal control parameters obtained in both methods are same. Also from plot 7 it is clear the more no. of solutions exists in narrow range i.e. there values are very closer.

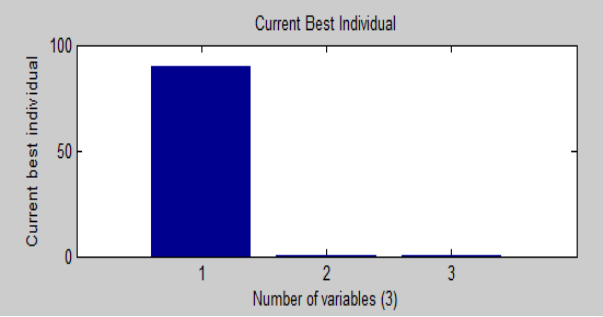


Fig.8 Current best individual versus Number of variables

Note: 1*=speed, 2*=depth of cut, 3*=feed

In Fig.8 current best individual versus number of variables plot shows that best individuals are at 90, 0.9 and 0.6 corresponding to the 1st, 2nd and 3rd variable. In this case these three variables i.e. v, DOC and f are found at their first level, third level and first level respectively.

At the 50th generation the objective function is obtained at function variables 1st level (cutting speed 90 m/min) 3rd level of depth of cut (0.6 mm) and 1st level of feed (0.09 mm/rev).

6. RESULTS AND DISCUSSIONS:

The experimental results were analyzed by using the GA Method which used for identifying the factors significantly affecting the performance measures. In the Fig. 8 the graphical results are obtained using GA Method.

From the GA method the final points corresponding to the different levels of cutting parameters are identified as 90 m/min (v, cutting speed,) , 0.09 mm/rev (f, feed) and 0.6 mm (d, DOC) respectively and objective function obtained is 298.5.

The Fitness value versus Generation values plot shows that there are the different fitness values at different generations and the curve consists of blue dots showing the mean fitness values while on the other hand the line containing the small black dots representing the best fitness values which is 2985.

In Fig.8 after selection and crossover, now there is a new population full of individuals. It can be loop through all the alleles of all the individuals, and if that allele is selected for mutation, then it can either change it by a small amount or replace it with a new value. The plot shows that best individuals are at 90, 0.9 and 0.6 corresponding to the 1st, 2nd and 3rd variable. In this case these three variables i.e. cutting speed, depth of cut and feed are found at their first level, third level and first level respectively.

By producing a "child" solution (new combination of function variables i.e. cutting speed, feed, depth of cut) using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each child, and the process continues until at 50th generation, the objective function is obtained at function variables 1st level (cutting speed 90 m/min) 3rd level of depth of cut (0.6 mm) and 1st level of feed (0.09 mm/rev) with a new population of solutions of appropriate size is generated.

6.1 Results Comparison of Taguchi and GA method

After comparing the results obtained from Taguchi and GA Method it is found that the optimal control parameters found from the Taguchi method are same as were selected from the GA method as given in Table 7. Hence it is also confirmed that the 1st level of speed, 3rd level of depth of cut and 1st level of feed are optimal control factors for optimized machining temperature.

In Taguchi method the output temperature was found 297.8 while 298.5 obtained in GA method which is closer to the Taguchi's output temperature.

Table-7 Results Comparison of Taguchi and GA method

Method	Speed, m/min	Depth of cut, (mm)	Feed mm/rev	Outpt Temp. (°C)	Differ-ence
Taguchi Method	90 m/min	0.6mm	0.09	297.8	0.7 °C
GA Method	90 m/min	0.6mm	0.09	298.5	

7. CONCLUSIONS

After solving the solution obtained from the turning of AISI 1018 work piece with K Type thermocouple it is observed that the technique adopted in this research work was an effective technique to determine the optimized control factors for obtaining the reduced temperature. Hence conclusions have been drawn. The control factors which played a very prominent and significant role were v, DOC and f. Out of these parameters the speed has maximum role in increasing the surface temperature during machining. The order of effectiveness of control parameters on the temperature is speed > DOC >f respectively. Feed has the least influence in rising the machining temperature during turning operation. temperature. Also the contribution of error was not significant and it is very low explaining that the speed, feed and depth of cut are major contributing factors. The experimentally obtained results were very accurate explaining that Taguchi technique adopted in this research work was an effective technique to determine the optimized control factors for obtaining the reduced temperature.

The results obtained from the GA method showed that the optimal parameters for cutting temperatures are cutting speed (1st level, 90 m/min) ,feed (1st level,0.09 mm/rev) and depth of cut at 3rd level i.e. 0.6 mm. Same values of control factors were identified in GA method. Hence it is confirmed that the 1st level of speed, 3rd level of depth of cut and 1st level of feed are optimal control factors for optimizing machining temperature in this experiment.

After comparing the results it is found that the optimal parameters found from the Taguchi method are very close as were identified from the GA method. In Taguchi method the output temperature were found 297.8 while 298.5 in GA method. The relationship was established b/w control factors v, f, DOC and formulated by regression equation which is verified by Genetic algorithm method that can be further used to find the optimum results.

REFERENCES

[1] M. Cotterell, E. Ares, J. Yanes, F. López, P. Hernandez, and G. Peláez, "Temperature and strain measurement during chip formation in orthogonal cutting conditions applied to Ti-6Al-4V," Procedia Engineering, vol. 63, pp. 922-930, 2013.

[2] B. Davoodi and H. Hosseinzadeh, "A new method for heat measurement during high speed machining," Measurement, vol. 45, pp. 2135-2140, 2012.

[3] A. Fata, "Temperature Measurement During Machining Depending On Cutting Conditions," P&A Sc and Tech., v01i2, pp. 16-21, 2011.

[4] P. J. Arrazola and T. Özel, "Investigations on the effects of friction modeling in finite element simulation of machining," International Journal of Mechanical Sciences, vol. 52, pp. 31-42, 2010.

[5] A. Fata, M. Bagheri, and P. Mottaghizadeh, "Tool Temperature Prediction during Machining by FEM with Experimental Validation," 2012.

[6] Y. Huang and S. Y. Liang, "Cutting temperature modeling based on non-uniform heat intensity and partition ratio," Machining science and technology, vol. 9, pp. 301-323, 2005.

[7] S. Zhang and Z. Liu, "An analytical model for transient temperature distributions in coated carbide cutting tools," International Communications in Heat and Mass Transfer, vol. 35, pp. 1311-1315, 2008.

[8] N. Aslan, "Application of response surface methodology and central composite rotatable design for modeling and optimization of a multi-gravity separator for chromite concentration," Powder Technology, vol. 185, pp. 80-86, 2008.

[9] S. Kalpakjian, Manufacturing engineering and technology: Pearson Education India, 2001.

[10] A. Muñoz-Sánchez, J. Canteli, J. Cantero, and M. Miguélez, "Numerical analysis of the tool wear effect in the machining induced residual stresses," Simulation Modelling Practice and Theory, vol. 19, pp. 872-886, 2011.

[11] G. Sutter, L. Faure, A. Molinari, N. Ranc, and V. Pina, "An experimental technique for the measurement of temperature fields for the orthogonal cutting in high speed machining," International Journal of Machine Tools and Manufacture, vol. 43, pp. 671-678, 2003.

[12] R. Doppalapudi, "Design-for-manufacturability (DFM) for system-in-package (SiP) applications," 2008.

[13] E. Kuljanic and M. Sortino, "Some Approaches in the Machining Research," in AMST'05 Advanced Manufacturing Systems and Technology, ed: Springer, 2005, pp. 41-55.

[14] K. Vernaza-Peña, J. Mason, and M. Li, "Temperature fields in aluminum during orthogonal cutting under different rake angles," in 2002 SEM International Congress and Exposition on Experimental and Applied Mechanics, 2002, pp. 10-12.

[15] Y. Karpat and T. Özel, "Analytical and thermal modeling of high-speed machining with chamfered tools," Journal of Manufacturing Science and Engineering, vol. 130, p. 011001, 2008.

[16] D. O'sullivan and M. Cotterell, "Temperature measurement in single point turning," Journal of materials processing technology, vol. 118, pp. 301-308, 2001.

[17] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira, "Response surface methodology (RSM) as a tool for optimization in analytical chemistry," Talanta, vol. 76, pp. 965-977, 2008.

[18] R. Komanduri and Z. Hou, "A review of the experimental techniques for the measurement of heat and temperatures generated in some manufacturing processes and tribology," Tribology International, vol. 34, pp. 653-682, 2001.

[19] Y. K. Potdar and A. Zehnder, "Temperature and deformation measurements in transient metal cutting," Experimental mechanics, vol. 44, pp. 1-9, 2004.

[20] P. C. Onyechi, B. S. Oluwadare, and N. Obuka, "Analytical Modeling of Temperature Distribution in Metal Cutting: Finite Element Approach."

[21] N. Abukhshim, P. Mativenga, and M. Sheikh, "Heat generation and temperature prediction in metal cutting: A review and implications for high speed machining," International Journal of Machine Tools and Manufacture, vol. 46, pp. 782-800, 2006.

[22] M. Sekulić, M. Gostimirović, P. Kovač, B. Savković, and Z. Jurković, "OPTIMIZATION OF CUTTING PARAMETERS BASED ON TOOLCHIP INTERFACE TEMPERATURE IN TURNING PROCESS USING TAGUCHI'S METHOD," Trends in the Development of Machinery and Associated Technology, 2011.

[23] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, Response surface methodology: process and product optimization using designed experiments vol. 705: John Wiley & Sons, 2009.

[24] F. Pusavec, P. Krajnik, and J. Kopac, "Transitioning to sustainable production–Part I: application on machining technologies," Journal of Cleaner Production, vol. 18, pp. 174-184, 2010.

[25] M. Rajemi, P. Mativenga, and A. Aramcharoen, "Sustainable machining: selection of optimum turning conditions based on minimum energy considerations," Journal of Cleaner Production, vol. 18, pp. 1059-1065, 2010.

[26] W. Yang and Y. Tarng, "Design optimization of cutting parameters for turning operations based on the Taguchi method," Journal of Materials Processing Technology, vol. 84, pp. 122-129, 1998.

[27] M. E. Capella-Peiró, A. Bossi, and J. Esteve-Romero, "Optimization by factorial design of a capillary zone electrophoresis method for the simultaneous separation of antihistamines," Analytical biochemistry, vol. 352, pp. 41-49, 2006.

ASSESSMENT OF HOSPITAL WASTE MANAGEMENT SYSTEM AND GENERATION RATES AT SMBBMU HOSPITAL LARKANA

Kishan Chand Mukwana*, Kamran Ahmed Samo**, Abdul Qayoom Jakhrani*

ABSTRACT

The rationale of this study was to assess the hospital waste management system at different medical wards and to determine the total daily waste generation rates at Shaheed Mohatrama Benazir Bhutto Medical University (SMBBMU) Hospital Larkana. For that purpose a detailed survey of hospital was conducted through a questionnaire, and personal interviews from selected staff and public. The acquired data was documented for waste generation, segregation, collection procedures, type of containers and storage, onsite and off-site transport, treatment of hospital waste and final disposal options. The number of beds, bed occupancy rate and number of operations performed during the year 2011 were also recorded. It was found that the average generation rate of hospital and infectious waste was about 1.072 and 0.453 kg/bed/day respectively. The maximum amount of infectious waste was recorded in Surgical-III ward with 49% and minimum amount was reported in ENT with 26% of the total hospital waste. It was discovered from the study that the existing hospital waste management system was lower than standards. The poor waste management system in the hospital was mainly due to ineffective segregation, collection, storage and dumping of waste. In addition, lack of training and protective equipment for sanitary staff, scarcity of funds and indefinite policy of waste treatment were also the main causes of inappropriate waste management. It was concluded from the study that clear lines of responsibilities and proper management system could be supportive for up-gradation of public health as well as hygienic conditions in the hospital.

Keywords: *Hospital waste, infectious waste, non-infectious waste, waste management system, management practices*

1. INTRODUCTION

The waste generated from hospitals is now acknowledged as a serious issue. It may have harmful effects either on the environment or on human beings through direct or indirect contact [1]. In developed countries, the various technologies and ways are performed for collection, storage, transport, and safe and secure disposal of hospital waste [2]. However, in developing countries, hospital waste has not received sufficient attention. Therefore, hazardous hospital waste is still handled and disposed off together with domestic wastes, thus creating a great health risk to municipal workers, the public, and the environment [3]. No doubt, the society is continuously trying to improve the health standards by establishing medical institutions [4]. These institutes can play a significant role to restore the community health by means of modern technologies. Medical institutes or establishments include hospitals, clinics, medical centers, private practices, home health care, blood banks, veterinary offices, research and clinical laboratories, and all licensed and unlicensed medical facilities [5]. In spite of the consideration given to hospital waste by the general public and relevant government bodies, the terms like health facility waste, hospital waste, medical waste, regulated medical waste, and infectious waste remain inadequately defined. No standard definition is universally accepted for these terminologies. Therefore, these terms are used synonymously.

Hospital waste is a solid or semi-solid waste produced in human health treatment giving facilities in the course of diagnosis or treatment of humans or animals. It includes pathogenic (infectious) and non pathogenic waste such as hazardous wastes, sharps, chemicals, pharmaceuticals, radioactive wastes, pressurized containers or cylinders. Non-infectious waste from healthcare settings may be regarded as similar to household waste and can be disposed off in a similar manner like municipal waste [6]. Infectious waste is one of the subset of hospital waste, that is able to transmit an infectious disease [7]. Infectious waste can also be further subdivided into various types such as micro-organisms, pathological tissue, blood product, animal body, surgery, laboratory, dialysis waste, isolation waste, blood contaminated, and fluid contaminated [8]. Infectious waste has to be managed in the proper way in order to minimize risk to the public health and environment.

The composition of hospital waste varies by the area, type and scale of medical facilities, clinic specialty and practice procedures [9]. The waste generation rates may

* Department of Energy and Environment Engineering, Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah. Email: mukwana_99@yahoo.com, aqunimas@hotmail.com

** Department of Electrical Engineering, Quaid-e-Awam University College of Engineering, Science and Technology (QUECST), Larkana. Email: kamransamo2@gmail.com

range from 0.25-7.0 kg/bed/day in seven European countries and the United States [10], 0.4-5.5 kg/patient/day in 12 developing and developed countries [11], and 0.11-3.9 kg/bed/day at hospitals of Japan, Turkey, US, Canada, India, Thailand, and Bangladesh [12]. In Jordon, public hospitals had a higher average waste production rate (6.10 kg/patient/day or 3.41 kg/bed/day) than private hospitals (4.02 kg/patient/day or 1.88 kg/bed/day) [13]. Unfortunately, the information on hospital waste generation rates and management system is quite rare in Pakistan except a few advanced hospitals. The amount of hospital waste and the risks to waste handlers can be reduced effectively with proper waste handling, segregation and resource recycling [14]. However, many factors influence the hospital waste management system and often linked to one another, which require a comprehensive study to determine the influence of each factor in the system. An in depth survey on the hospital waste generation methods can be informative and beneficial for planning and enhancement of waste management system. This study investigated the generation rates of hospital waste in Shaheed Mohatrama Benazir Bhutto Medical University (SMBBMU) Teaching Hospital Larkana. The findings of this study could be useful for authorities to improve the methods of collection, handling and disposing off hospital waste. Consequently, it will also support to decrease the quantity of infectious waste by means of categorization and separation methods of waste at the source.

2. MATERIALS AND METHODS

2.1. Study Area

Larkana is fourth largest city of Sindh province. It is located in the north-western part of Sindh. It lies on $27^{o}33'$ N Latitude and $68^{o}12'$ E Longitude. Shaheed Mohtarma Benazir Bhutto Medical University (SMBBMU) has started functioning at Chandka Medical College on 12th April 2008. SMBBMU Hospital comprises four separate wings, namely Chandka Teaching Block, City Block, Paeds Block and Sheikh Zaid Women Hospital. SMBBMU is the leading referral medical center in northern part of Sindh and neighboring areas of Balochistan Province. However, till today no environment friendly waste management strategy had been incorporated in the hospital [15].

2.2. Data Collection and Analysis

A study was conducted to collect the data regarding the number of wards, units, beds and patients. The collection of hospital waste samples and quantitative analysis were conducted in 2011. Initially, a trial collection exercise of one week was adopted before initiation of the regular collection program. The purpose of this trial exercise was to identify the main waste types and characteristics, determine their respective quantities and plan the regular collection and segregation of waste. In all cases, waste collection and measurement took place for six working weeks, one week per month in the two collection periods. Different color coded waste containers, plastic polyethylene bags, an electronic balance, calculator and recording forms were used for waste handling and recording. The color coded waste containers were labeled with the international sign of hospital waste. Their capacity was about 03 liters. The containers were washed and reused every single day. Plastic polyethylene bags were purchased from a local market and were used for waste dumping after weighing. The plastic bags were suitable for preventing liquid leakage and leachate of hospital waste. An electronic balance with accuracy up to a tenth of a gram was used for weighing purpose. Daily waste recording forms were used for computational work and documentation of generated waste.

Collection containers were kept at suitable locations in each ward, laboratory and department of the hospital. Hospital personnel were supervised that the waste is to be dropped in the proper color coded containers. All wastes dropped in containers were weighed on a daily basis. The information regarding number of patients, occupied beds as well as medical examinations were recorded in computerized Microsoft excel spreadsheets. Furthermore, a survey was carried out through questionnaire, personal interviews of staff and patients, observations, and official record of the hospital for assessment of existing hospital waste management system for validation of public views and visual observations. The responses and documentations received from survey were reviewed for completeness and consistency before statistical analyses. The hospital waste was separated into different categories such as paper, textiles, plastic, glass, food residues, rubber and metal for determination of their type and weight percentage. Moreover, the waste generation rate was computed by dividing the total weight of waste (in kg) generated per day with the number of beds in the hospital (the empty bed was not count in the calculation) expressed as kg/bed/day or dividing the total weight of waste (in kg) generated per day with number of patients attended daily in the hospital (inpatient and outpatients) expressed as kg/patient/day [16].

2.3. Existing Hospital Waste Management System at SMBBMU Hospital

The existing waste management system at SMBBMU Hospital was comprised of five elements namely generation, onsite handling, storage and processing, collection, transport and disposal. Generation type was consisted of both infectious and non-infectious wastes. Hospital waste generated close to the bedside usually thrown into the small waste bins with no lid or cover,

which produced unhygienic environment near the bed. Hospital waste was neither segregated nor it was color coded for separation of different types of wastes. All types of waste were stored in the containers whether it was infectious or non-infectious. The municipal workers handling the waste normally empties the smaller containers into the larger ones placed in the corridors. No suitable isolated storage area was designated in the hospital. Protective devices were also not given to the waste handling staff for safety. The sanitary staff was just handling the waste as normal domestic waste. The generated hospital waste was usually collected in open trolleys and waste bins. This practice is supposed to create health hazards as the vectors in the air came into the contact with the contaminated waste that later on became source of disease for public and employees of the hospital. Furthermore, the waste collected was transferred from inside of the hospital either on shoulders or by means of manual carts, with the possibility of infectious waste dropping on walking ways. No municipal container was placed in the front road of SMBBMU city hospital.

The waste placed in municipal containers contained the waste of SMBBMU as well as other adjoining areas. The waste from the containers was regularly picked up and shifted by the municipal authorities to the final destination, which was an open dump site. Consequently, the hospital waste mixed up with municipal waste at the final disposal site. The scavengers were also seen outside the hospital and municipal dumping area, searching for recyclable materials. The recyclable materials especially syringes returned to the market somehow after being washed. Such syringes contained pathogens, leads to very harmful consequences if used to patients. Although, Government of Sindh provided an incinerator for treatment of SMBBMU Hospital waste, but it was not operating due to unavailability of gas connection and other allied facilities.

3. RESULTS AND DISCUSSIONS

The capacity and cause of hospital waste created in the wards was evaluated by means of Statistical Package for the Social Sciences (SPSS) software. It was found that the average generation rate of hospital waste was 1.072 kg/bed/day and infectious waste was 0.453 kg/bed/day. The highest amount of infectious waste was generated in Surgical-III ward with 49%, Urology 44%, Medical-I 41% of total hospital waste. Minimum amount of infectious waste was reported in ENT with 26% and Gynecology I-II 29% of the total generated waste.

Outdoor patients per day, bed strength, average bed occupancy rates and total number of operations performed during the year 2011 are shown in Figures 1 to 4. There were a total of sixteen number Outdoor Patients Department (OPD) sections in the hospital as shown in Figure 1. The hospital provided free of cost facilities to all patients. The number of OPD patients sometimes crossed beyond 3500 daily. However, the average number of OPD patients per day for the year 2011 stood at 3040. The average maximum number of patients per day in OPD was recorded in Pediatrics Medicine with 350 and the minimum patients were reported in Neuro Surgery with 50. The number of patients was 300 each in General Medicine and Gwynne Obis. A total of 250 patients were came for diagnosis and treatment purpose in General Surgery, Cardiology and Urology, whereas 200 patients came for consultation in EYE and Orthopedics, 150 in ENT, Dental and Chest and 140 in Dermatology.

Moreover, there were about 25 different wards and units in SMBBMU Hospital. The bed strength of different wards and units is graphically shown in Figure 2. The total number of beds in the hospital was stood at 1356. Out of these beds 725 were reserved for male and 631 for female patients. The total maximum number of bed strength for male and female was 200 in Pediatrics Medicine and in Gynecology wards respectively. The maximum number of male bed strength was in Pediatrics Medicine with 165 beds and female bed strength Gynecology Unit I&II with 200 beds respectively. The second maximum number of beds strength for male was in Medicine Unit-II with 60 beds. The second and third maximum beds strength for female was found in Pediatrics Medicine with 35 beds and Medicine Unit-II and Eye wards with 30 beds respectively. The minimum number of beds for both male and female were in critical care unit (CCU) with two beds only. The second and third minimum number of beds strength for male as well as female was noted in Peng ward and Dermatology with four and five beds and three and five beds respectively.

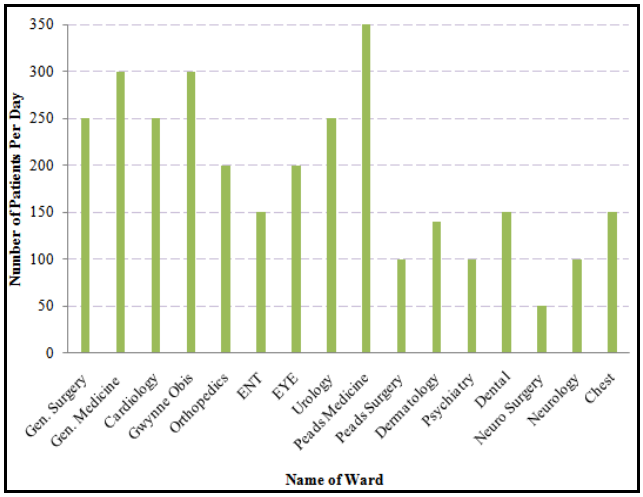


Figure 1. Ward-wise Number of Outdoor Patients at SMBBMU Hospital in 2011

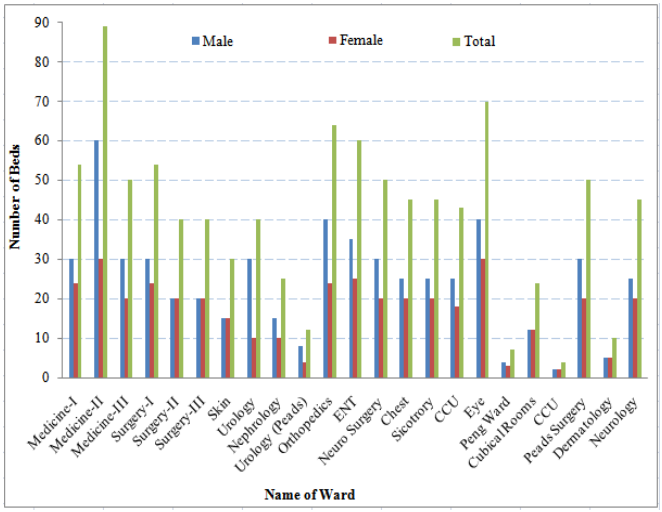


Figure 2. Ward/Unit-wise Bed Strength of SMBBMU Hospital in 2011

The average bed occupancy for the Year 2011 in SMBBMU Hospital is demonstrated in Figure 3. The total numbers of occupied beds were found to be 728 (54% of the total beds) out of 1356. The maximum beds occupancy was noted in Medicine Unit-I with 94.4% and the minimum was found in Pediatrics Medicine ward with 19%. Second and third most occupied beds were noted in Urology with 80% and Surgical-III with 78%. Similarly, second and third least occupied beds were found in Surgery Unit-II with 28% and Gynecology Unit I&II with 35%. In addition, the total number of major and minor operations performed during the year 2011 was 57681 with a monthly average of 4804 operations as shown in Figure 4. The average monthly maximum operations were made in the month of September with 5707 and minimum operations in the month of May with 3478.

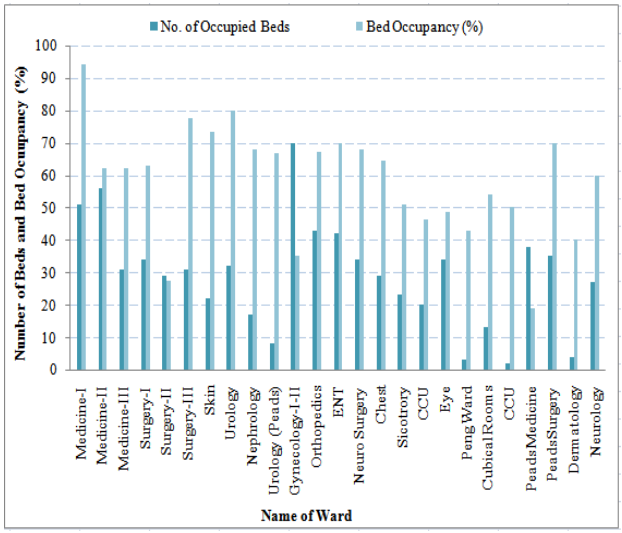


Figure 3. Ward/Unit-wise Bed Numbers and Average Bed Occupancy of SMBBMU Hospital in 2011

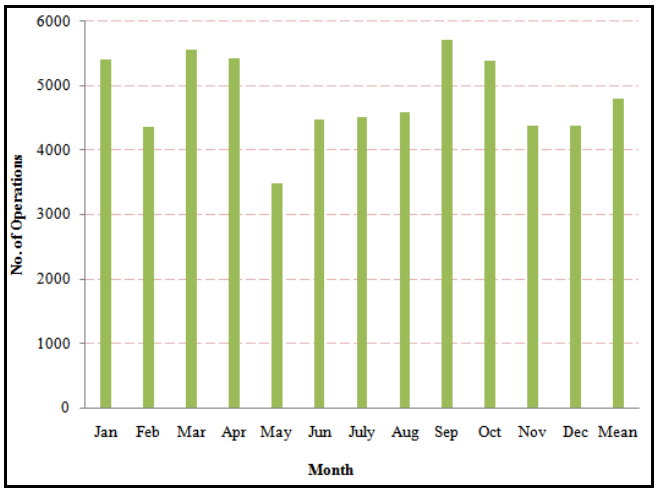


Figure 4. Number of Major and Minor Operations Performed at SMBBMU Hospital in 2011

The total generated waste was determined by examining the bed strengths on daily basis during study period. The total weight of hospital waste during study period was found to be 781 kg and the infectious waste were 330.4 kg as shown in Table 1. The component of infectious waste in the wards, OPD, X-Rays, Radiology and Research Lab was around 37 % of the total generated waste. The share of infectious waste from the total generated waste was found to be 100% in operation theaters, blood bank and pathological laboratories. Because the waste generated by such wards and units is considered to be pathogenic.

Table 1. Weight of Generated Hospital Waste and Infectious Waste

Sr. No.	Name of Ward and Unit	Total Waste (kg)	Infectious Waste (kg)
01	Wards	661.2	248.7
02	OPD, X-Rays, Radiology and Research Lab	53.8	20.3
03	Operation Theaters	24.2	24.2
04	Blood Bank & Pathology lab	16.0	16.0
05	Plastic drips/vials/ampoules	25.8	21.2
Total		**781.0**	**330.4**

4. PROPOSED MANAGEMENT SYSTEM FOR HOSPITAL WASTE

The prevailing situation in the investigated hospitals was not attractive. It was observed that hospital waste was not properly managed. The infectious waste was generated from the hospital was composed of blood as well as blood socked substances, body parts, body fluids, bandages, glucose bags, syringes etc. The generated hospital waste was treated as ordinary domestic municipal waste and

dumped off in vacant areas of city. The scavengers were found as main collectors at the curbside and at the final disposal points. The collected reusable waste materials were marketed by the scavengers to fulfill their economic needs. However, such activities could pose serious threats due to possible chance of contracting and spreading infectious diseases.

4.1. Proposed Management Plan

The SMBBMUH is biggest hospital in the northern part of Sindh with more than 25 various wards. The SMBBMUH has 1356 beds with a bed occupancy rate of about 54%. Therefore, a strategy of total quality management (TQM) was proposed to be adopted to reduce generation of waste in the SMBBMUH. This perception was first adopted in industrial sites. It is a valuable tool and is able to enhance and improve the health and hygienic conditions in the hospitals [17]. This approach could be applied by using the acronym FOCUSPDCA which is explained as under:

F = Find a process for improvement

O = Organizing trained team for the process

C = Clarify and simplify the process

U= Understand sources of variation

S = Select the improvement

P = Plan the improvement

D = Doing the improvement

C = Check and study the outcomes

A = Act to achieve the gains

The waste management plan must be composed of different members headed by Medical Superintendent. The members includes Heads of every department, Infection Control Officer, Chief pharmacist, Radiation Officer, Senior Nursing Officer, Hospital manager, Health and safety Engineer, Financial Controller and waste management Officer. These officials should be responsible to ensure the implementation of waste management plan by relevant staff.

4.2. Strategy for Storage, Handling and Disposal of Waste

An implementation approach was formulated for handling, storage and final disposal of hospital waste. This approach includes waste handling method, color coding, design and placement of containers and time table of waste collection. Waste handling method consists of different measures like, trained staff to put the waste into color coded bins. Sanitary staff must wear gloves and masks while handling the waste and also wear complete protective clothing when required. The waste containers must be replaced when full and immediately place a new container when transporting the previous container. Color coded containers facilitates the waste generators to differentiate infectious and non-infectious waste. It is recommended that infectious waste and sharps must be placed in yellow color, and non-infectious waste in black color containers. Furthermore, infectious waste must be placed in closed pedal type containers and sharps in strong puncture proof containers. The design of containers may vary from one type of waste to other type. However, it is proposed that the containers for sharps and non-infectious type wastes should be approximately 1.5'x1.5'dia x 2'deep in size. However, the carrying capacity for sharp containers should be one kilogram and the pedal type containers should be of 10 kilograms. The pedal type containers should be placed according to the ward-wise bed occupancy rates for collection of infectious waste. It was also suggested that the hospital waste must be collected two times per day, the first one at 08:00 am and the last one around 08:00 pm. The infectious waste collected from hospital must be stored in biohazard room and it should be transported early in the morning towards treatment facilities. Non-infectious waste may be sent to the municipal containers for final disposal.

5. CONCLUSIONS AND RECOMMENDATIONS

Surveys are the valuable quality improvement tools in all hospitals for enhancement of public health and protection of environment. It was revealed from the survey that the hospital waste not properly managed as per standards. The infectious and non-infectious waste was collected, transported and disposed off along with municipal waste. The waste containers were neither clean nor disinfected after use. The sanitary staff was not properly trained to handle the hazardous waste. The collected waste was carried away in open containers on hand carts and sometimes on shoulders to its final disposal points. No proper waste management plan was available and implemented. The study revealed that the average generation rate of hospital waste was 1.072 kg/ bed/day and the infectious waste was 0.453 kg/bed/day (42% of the total hospital waste). The highest amount of infectious waste was generated in Surgical-III ward with 49% and the minimum quantity was noted in ENT with 26% of the total generated waste. An effective medical waste management program was found to be indispensable in all levels of hospital units. The authors suggest the following measures to get meaningful outcomes:

a) Different types of color coded containers should be used for infectious and non-infectious waste.

b) The waste bins should be properly placed, handled, timely collected and transported.

c) Protective clothing and materials as well as training should be given to the sanitary staff.

d) Need of public awareness.

e) Proposed management plan should be implemented with true sprit to reduce the chance of spreading infectious diseases.

ACKNOWLEDGEMENT:

The authors acknowledge the cooperation extended by the management of Shaheed Mohtarma Benazir Bhutto Medical University (SMBBMU) hospital.

REFERENCES

[1] El-Salam, M.M.A., 2010. Hospital waste management in El-Beheira Governorate, Egypt. Journal of Environmental Management 91, 618–629.

[2] Tudor, T.L., Noonan, C.L., Jenkin, L.E.T., 2005. Health care waste management: a case study from the National Health Service in Cornwall, United Kingdom. Waste Management 25 (6), 606–615.

[3] Silva, C.E., Hoppe, A.E., Ravanello, M.M., Mello, N., 2005. Medical waste management in the south of Brazil. Waste Management 25 (6), 600–605.

[4] Chaerul, M., Tanaka, M., Shekdar, A.V., 2008. A system dynamics approach for hospital waste management. Waste Management 28, 442–449.

[5] Labib, O.A., Hussein, A.H., El-Shall, W.I., Zakaria, A.O., Mohamed, M.G., 2005. Evaluation of medical waste incinerators in Alexandria. Journal of the Egypt Public Health Association 80, 390–404.

[6] Rushbrook, P., Zghondi, R., 2005. Better Health Care Waste Management. World Health Organization, Regional Office for the Eastern Mediterranean Regional Center for Environmental Health Activities (CEHA), Jordan, Amman.

[7] Huang Mei-Chuan, Lin J.J., 2008. Characteristics and management of infectious industrial waste in Taiwan. Waste Management 28, 2220–2228.

[8] Environmental Protection Administration (EPA), 2010. Medical Waste Information Network Taipei, Taiwan. Available at wm.epa.gov.tw/medicalwaste/Contents/J03.html. Accessed 28.06.2011.

[9] Cheng, Y.W., Li, K.-C., Sung, F.C., 2010. Medical waste generation in selected clinical facilities in Taiwan. Waste Management 30, 1690–1695.

[10] Liberti, L., Tursi, A., Costantino, N., Ferrara, L., Nuzzo, G., 1996. Optimization of infectious hospital waste management in Italy Part II. Waste characterization by origin. Waste Management and Research 14, 417–431.

[11] Chung, S.S., Lo, C.W.H., 2003. Evaluating sustainability in waste management: the case of construction and demolition, chemical and clinical wastes in Hong Kong. Resources, Conservation and Recycling 37, 119–145.

[12] Mohee, R., 2005. Medical wastes characterisation in healthcare institutions in Mauritius. Waste Management 25, 575–581.

[13] Bdour, A., Altrabsheh, B., Hadadin, N., Al-Shareif, M., 2007. Assessment of medical wastes management practice: a case study of the northern part of Jordan. Waste Management 27, 46–759.

[14] Sabour, M.R., Mohamedifard, A., Kamalan, H., 2007. A mathematical model to predict the composition and generation of hospital wastes in Iran. Waste Management 27, 584–587.

[15] Samo, K. A., 2012. Assessment of Infectious Waste of Hospitals in Larkana City. Master's Thesis, Department of Energy & Environment Engineering, Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah.

[16] Kagonji, I.S., Manyele, S.V., 2011. Analysis of the measured medical waste generation rate in Tanzanian district hospitals using statistical methods. African Journal of Environmental Science and Technology 5 (10), 815-833.

[17] Rad, A.M., 2005. A survey of total quality management in Iran. Int. J. Health Care Qual. Assur. Inc. Leadersh. Health Serv. 18 (4–5), xii–xxxiv.

STUDY OF ENERGY POTENTIAL FROM MUNICIPAL SOLID WASTE OF MIRPURKHAS CITY

Kishan Chand Mukwana*, Saleem Raza Samo*, Muhammad Mureed Tunio*, Abdul Qayoom Jakhrani*, Muhammad Ramzan Luhur**

ABSTRACT

The purpose of this study was to determine the energy potential of combustible portion of total generated municipal solid waste of Mirpurkhas city. Firstly, the survey was conducted for selection of suitable locations based on types of areas (residential, open areas and commercial units), living standard, occupied area and number of people in a single family. Then, the total collected waste was separated manually with the help of sanitary workers for characterization and quantification. Finally, the energy potential was estimated from combustible part of generated waste. It was found from the study that the total municipal solid waste generated from the city was about 200 tons per day. The share of paper waste was about 20% and wood waste was 15% of the total generated waste. The study revealed that 60 to 70% of the total municipal waste was the combustible waste, which can produce 130 million kcal energy (heat) per day. The produced energy can be used for industrial processes in sugar mills, cement kilns or brick kilns, instead of throwing it uselessly in open areas. The open dumping of waste was the main cause of environmental pollution in the city.

Keywords: *Waste management, energy potential, combustible waste, industrial processes*

1. INTRODUCTION

Municipal solid waste is an undesirable material discarded by human activities. It mostly consists of solid, semisolid or liquid materials, thrown away from households, commercial and industrial areas. Solid waste generation and its consequences for the public and the surroundings are global issues. The complications of the physical and chemical characteristics of solid waste are a challenge for waste managers, especially in developing countries. It requires a vibrant and effective policy for waste management and proper legislation [1-3]. Besides huge complications, municipal solid waste (MSW) contains a substantial amount of energy, which can be utilized efficiently for production of heat and electricity [4-5]. Usually, the higher heating value of MSW is between 18 and 20 GJ tonne^{-1}, and the lower heating value is between 8 and 12 GJ tonne^{-1}. Since, the bituminous coal has lower heating value of 23.9 GJ tonne^{-1}. It is clear from the statistics that the lower heating value of MSW is 42 percentage of the bituminous coal. It could be economically feasible to utilize the energy of discarded waste in environment friendly manner instead of throwing it without energy recovery [6]. The changing of living standards, decreasing utilization of un-disposable materials and unnecessary packing of different items are responsible for increasing the waste generation quantity. The problems associated with MSW management are complicated as the characteristics and composition of waste is different and there are also financial limitations on public services. It is not only the issue of land but also air and water as it pollutes every environment [7].

It was reported that the solid waste generation rate ranges from 0.283 to 0.612 kg/capita/day and the growth rate was 2.4% per year in Pakistan [8-10]. It was observed that the municipal solid waste was typically dumped over the low lying areas. The solid waste is openly burned to reduce its volume and lengthen the life span of the dumpsite. However, the dumped waste is not completely burned as it contains moisture, which produces unwanted smoke. That results unaesthetic look and unsanitary conditions in the area [11-13]. However, that land could be used for the production of more valuable things instead of dumping site for waste materials. Since, the potentially valuable recyclable materials are lost due to direct burning. Moreover, no separation of waste is being carried out for different types of wastes and the population is also unaware about the proper ways and means of waste disposal system. Consequently, direct disposing and burning of waste results environmental and public health problems in the area [14-15].

2. MATERIALS AND METHODS

2.2 Study Area

Mirpurkhas city was selected for the study of energy potential of municipal solid waste. It is the fourth largest city of Sindh province with a population of 0.4 million. The site survey was carried out in 2012, for selection of different locations in the city area. The types of areas included commercial and residential colonies. Different

* Energy & Environment Engineering Department, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah.

** Mechanical Engineering Department, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah.

criteria were considered for the selection of locations, such as type of commercial unit, living standard, occupied area and number of people in a single family. The total collected waste was separated manually with the help of sanitary workers by means of providing different plastic bags. After the separation of waste, it was weighed and quantified. The heating value of solid waste from residential, open areas and commercial units is 4300 Btu/lb, 6500 Btu/lb and 2500 Btu/lb respectively. The efficiency of solid waste combustion depends on the waste characteristics, ignition temperature and incineration technologies. The combustion efficiency is commonly ranges from 50% to 85%. Since, 70% efficiency is being considered for complete combustion of waste through incinerator [16-17].

2.3 Estimation of Energy Potential from Solid Waste

The following equation was used for solid waste generation on hourly basis from daily data [17]:

$$\text{Total of average solid waste generation per hour } (lb/h) = \frac{\text{Average of solid waste per capita (kg/day) x 2.2046}}{\text{24 hour}}$$

The energy potential from solid waste was calculated as [17]:

$$\textit{Energy (kWh)} = \frac{\text{Total of average solid waste generation/hour} \times \text{heat value} \times \text{efficiency}}{\text{3412 Btu per hour}}$$

The value of 3412 Btu/hour is being used to obtain value in electrical energy produced that is kWh and 1 kWh equal to 3412 Btu/hour or 1.341 hp.

3. RESULTS & DISCUSSIONS

It was found from the study that the municipal solid waste was composed of heterogeneous mixtures. It was openly thrown away in low lying areas of the city. The solid waste was categorized based on different constituents as shown in Table 01.

Table 01: Solid Waste category & constituents of Mirpurkhas Solid Waste [18-20]

S. No.	Category of Solid Waste	Waste constituents
1	High-grade paper	Office and computer paper
2	Mixed Paper	Mixed colored papers, magazines, glossy paper, and other paper and newsprint Mixed colored papers, magazines, glossy paper, and other paper and newsprint
3	Newsprint	Newspaper
4	Yard Waste	Branches, twigs, leaves, grass and other plant material
5	Food Waste	All food waste excluding bones
6	Glass	Clear and colored glass
7	Plastics	All types of plastics
8	Ferrous metals	Iron, steel, and tin and metal cans
9	Non-ferrous metals	Primarily Aluminum, Aluminum cans, copper, brass and lead
10	Wood	Timber, wood products, pallets and furniture
11	Rubber	Tyres, footwear, wire cords, gaskets
12	Textiles	Furniture, clothing, and footwear
13	Miscellaneous	Other organic and inorganic materials, including rock, sand, dirt, ceramics, plaster, bones ashes, etc.

The total quantity and the collected solid waste results are given in Table 02. It was found from the study that about 200 tons of solid waste was being generated from Mirpurkhas city. Out of this total generated waste about fifty percent was being collected by the municipal staff (sanitary staff), while the rest of the waste left out at the open areas. The municipal authorities have found four final disposal sites out of the city, where they carry out the collected solid waste and performed open dumping. The transport fleet comprised of thirty vehicles which were open from top and the solid waste was manually loaded over them by the sanitary workers.

Table 02: Total quantity of the collected solid waste

S. No.	Item	Volume/Quantity
1	Total Waste Generation	200 tons per day
2	Collection per day	100 tons per day
3	Left out waste	100 tons per day
4	No. of disposal sites	04 sites
5	No. of sanitary workers employed	246
6	No. of vehicles in use	30

It was found that paper waste was dominant substance of the generated solid waste stream, which was about 20 percent of the total. The share of plastic waste was about 20 percent. Since, the vegetable waste, kitchen waste and wood waste were 15 percent each, while the glass waste was about 5 percent of the total. The composition of the solid waste found in the generated solid waste is shown in Figure 1.

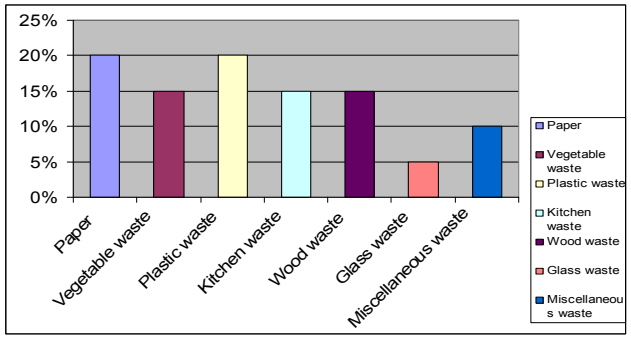


Figure 1: Composition of Solid Waste in Mirpurkhas

It was also found from the study that combustible solid waste was about 60 to 70% of the total generated waste. The energy or heat can be produced from that part of waste. The energy from combustible waste can be used by

sugar industries, brick kilns and cement factories in the vicinity of Mirpurkhas city. The study revealed that combustible waste contains approximately 130 million kcal energy (heat) per day, which can reduce the fuel consumption as well as environmental pollution in the area.

4. CONCLUSION

The total collected waste was separated manually with the help of workers for characterization and quantification. The energy potential was estimated from combustible part of generated waste by analytical methods. It was found that the total municipal solid waste generated from the city was about 200 tons per day. The share of paper waste was about 20% and wood waste was 15% of the total generated waste. It was discovered that 60 to 70% of the total municipal waste was the combustible waste, which can produce 130 million kcal energy (heat) per day. It is concluded that the produced energy from solid waste can be effectively used for heat and energy production in sugar mills, cement or brick kilns in the vicinity of the city, instead of dumping into open areas.

ACKNOWLEDGEMENT

The authors highly acknowledge the facilities provided by the authorities of Quaid-e-Awam University of Engineering Science and Technology Nawabshah and municipal administration of Murpurkhas city.

REFERENCES

[1] Environmental Protection Agency Government of Pakistan & JICA, (2005) "Solid Waste Management Guidelines" Islamabad, Pakistan.

[2] Farooq, M. K., & Kumar, S. (2013). "An assessment of renewable energy potential for electricity generation in Pakistan". Renewable and Sustainable Energy Reviews, 20, 240-254.

[3] Masood, M., Barlow, C. Y., & Wilson, D. C. (2014). "An assessment of the current municipal solid waste management system in Lahore, Pakistan. Waste Management & Research, 0734242X14545373.

[4] Habib, K., Schmidt, J. H., & Christensen, P. (2013). A historical perspective of Global Warming Potential from Municipal Solid Waste Management". Waste management (New York, NY), 33(9), 1926-1933.

[5] Bhutto, A. W., Bazmi, A. A., & Zahedi, G. (2012). "Greener energy: issues and challenges for Pakistan—solar energy prospective". Renewable and Sustainable Energy Reviews, 16(5), 2762-2780.

[6] Kathirvale, S., Muhd Yunus, M. N., Sopian, K., & Samsuddin, A. H. (2004). "Energy potential from municipal solid waste in Malaysia". Renewable energy, 29(4), 559-567.

[7] Noor, Z. Z., Yusuf, R. O., Abba, A. H., Abu Hassan, M. A., & Mohd Din, M. F. (2013). "An overview for energy recovery from municipal solid wastes (MSW) in Malaysia scenario". Renewable and Sustainable Energy Reviews, 20, 378-384.

[8] Magsi A, 2008. "Environmental Assessment of Solid Waste of Mirpurkhas city & its Management". Bachelor's Thesis, Department of Energy & Environment Engineering, Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah.

[9] Periathamby, A., Hamid, F. S., & Khidzir, K. (2009). "Evolution of solid waste management in Malaysia: impacts and implications of the solid waste bill, 2007". Journal of material cycles and waste management, 11(2), 96-103.

[10] Zurbrugg, C. (2002). "Urban solid waste management in low-income countries of Asia how to cope with the garbage crisis. Presented for: Scientific Committee on Problems of the Environment (SCOPE)" Urban Solid Waste Management Review Session, Durban, South Africa, 1-13.

[11] Wei, J. B., Herbell, J. D., & Zhang, S. (1997). "Solid waste disposal in China-Situation, problems and suggestions". Waste management & research, 15(6), 573-583.

[12] Dolgen, D., Sarptas, H., Alpaslan, N., & Kucukgul, O. (2005). "Energy potential of municipal solid wastes". Energy sources, 27(15), 1483-1492.

[13] Cheng, H., & Hu, Y. (2010). "Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China". Bio-resource technology, 101(11), 3816-3824.

[14] Arena, U. (2012). "Process and technological aspects of municipal solid waste gasification". A review. Waste management, 32(4), 625-639.

[15] Johari, A., Ahmed, S. I., Hashim, H., Alkali, H., & Ramli, M. (2012). "Economic and environmental benefits of landfill gas from municipal solid waste in Malaysia". Renewable and Sustainable Energy Reviews, 16(5), 2907-2912.

[16] Zahari, M. S., Ishak, W. M. F. W., & Samah, M. A. A. (2010). "Study on solid waste generation in Kuantan, Malaysia: its potential for energy generation". Int. Eng. Sci. Technol, 2(5), 1338-1344.

[17] Amber, I., Kulla, D. M., & Gukop, N. (2012). "Generation, characteristics and energy potential of solid municipal waste in Nigeria". Journal of Energy in Southern Africa, 23(3), 47.

[18] Reddy, P. J. (2011). "Municipal Solid Waste Management: Processing-Energy Recovery-Global Examples". CRC Press.

[19] Ludwig, C., Hellweg, S., & Stucki, S. (2002). "Municipal solid waste management". Springer-Verlag.

[20] Sharholy, M., Ahmad, K., Mahmood, G., & Trivedi, R. C. (2008). "Municipal solid waste management in Indian cities–A review". Waste management, 28(2), 459-467.

MODIFIED SPEED PROTOCOL FOR WIRELESS SENSOR NETWORKS

Irfana Memon*, Nisar Ahmed Memon**, Fozia Noureen**

ABSTRACT

Nowadays wireless sensor networks (WSNs) have grown tremendously because of the easy deployment. In WSNs, sensor nodes monitor the area, and send the desired information to a remote sink for user access. Different protocols are used to select the route for data transmission in WSNs. These protocols might be different based upon WSNs applications and its architecture. In some of WSNs applications, data should be reached at the destination within the delay bound time; otherwise data will not be acceptable. T. He et al [26] have proposed a protocol called SPEED with considering delivery time. The main objective of SPEED protocol is to provide communication within given time. SPEED routing protocol is basically based on a single hop delay for real time traffic and energy metrics has not been considered as a key parameter during routing path selection. Consequently energy depletion of selected nodes will be faster. In present work the SPEED protocol is redefined incorporating the energy parameter as an additional factor and the modified protocol showed better results in terms of energy efficiency.

Keywords: *WSNs, Data routing in WSNs, QoS in routing protocol, SPEED routing protocol*

1 INTRODUCTION

Wireless sensor networks (WSNs) have got a lot of consideration for use in real environmental monitoring WSN applications, including sea depth measuring, metrological hazard detection, earthquake alerting, fire detection and enemy detection in battle field; some applications are summarized in [1]. WSNs are collection of a large number of sensor nodes having capabilities such as computation, sensing, and wireless communications.

Sensor nodes have limited resources in terms of energy, computation capability, and limited communication range that have been thoroughly discussed in [2]. Almost all the application areas where WSNs are deployed need to sense events in their surroundings and report them to a remote base station (Sink). In WSNs, routing protocols are used for communication between sensor nodes, so that sensor nodes can transmit their information to the remote Sink. Many protocols for routing the information in WSNs have been proposed in the literature [3]; and their function depends upon the type of network structure or the network operations for a specific application model. Almost all of these protocols can be categorized in 3 classes based on structure of the network: flat routing protocol, hierarchical routing protocol, and location-based routing protocol.

Flat-based routing protocols: All sensor nodes perform same function/role in the network. The protocols proposed in [4]-[12] belonging in this category.

Hierarchical-based routing protocols: In this category, sensor nodes perform different tasks in the network. For example, in clustered sensor network, cluster head nodes do some processing (aggregation) on the data to save energy. The protocols proposed in [13]-[21] belonging in this category.

Location-based routing protocols: In this category, sensor node selects the route for data transmission according to the location in the network. These protocols use the location information to send the data to destination instead of the whole network. The protocols proposed in [22]-[25] belonging in this category.

Up to now, most of the existing routing protocols for WSNs consider communication overhead reduction for improving energy efficiency. However, in some applications, the main requirement is a desired Quality of Service (QoS) in terms of bandwidth, delay and information throughput. However, these requirements will affect the selection of routing protocol for a specific application. For example, military applications: data should be reached at the destination within a specified time period.

The protocol 'SPEED' proposed in [26] is considered as most popular QoS routing protocol for WSNs. An extension of SPEED protocol has been proposed by Lee et al. called MMSPEED [27]. The protocol MMSPEED improves the SPEED protocol [26]. The MMSPEED guarantee message delivery in given timeliness by using multiple speeds and adopts probabilistic multipath forwarding strategy for reliability demands of different applications. But, these protocols did not consider the

* Corresponding author

** Department of Computer Systems Engineering Quaid-e-Awam University of Science & Technology, Nawabshah.
irfanahameed@quest.edu.pk drnisar@quest.edu.pk engrnoureen@quest.edu.pk

energy consumption. Since the network lifetime mainly depends on the battery life of sensor nodes, it is important to consider energy consumption while satisfying QoS requirements. In this research work we present an energy efficient protocol called modified SPEED protocol which is developed by incorporating energy consideration in the existing SPEED protocol, resulting enhancement in the network lifetime. Simulation results bring out that our modified SPEED protocol outperforms SPEED protocol in term of energy efficiency and overall throughput.

The rest of the paper is organized as follows. In Section 2, we discuss SPEED routing protocol in details and proposed improvements. In Section 3, we present the performance evaluation of proposed modified-SPEED protocol, and the paper is concluded in the Section 4.

2 'SPEED' PROTOCOL [26]

This protocol is a spatiotemporal communication protocol. The key difference from the other geographic location routing protocols is that SPEED takes into account the delivery time.

In SPEED protocol, each node maintains its neighbor list having information about one- hop neighbors and selects the path for data transmission using geographic forwarding. SPEED ensures a specific speed for each packet, which is used by nodes to calculate delay bound for that packet. This delay is evaluated as follows:

delay = distance between source to destination / speed **.(1)**

In this case the delay bound is proportional to the distance from source to destination. Furthermore, in congested networks, SPEED prevents voids. Figure 1 illustrates SPEED protocol, redrawn from [26].

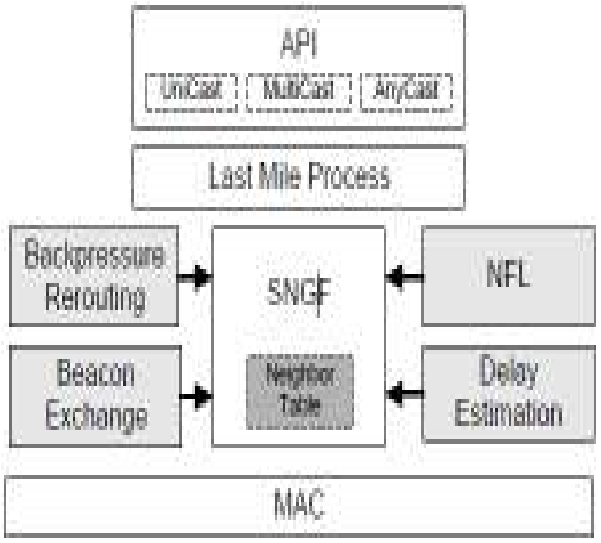


Figure 1: SPEED routing protocol [26]

There are four **API (application-levels):** (i) AreaMulti-castSend (ii) AreaAnyCast-Send (iii) Uni-castSend, and (iv) Speed-Receive.

The neighbor beacon exchange scheme provides information about its neighbors and their location. In SPEED, each node sends a beacon packet periodically to exchange information with its neighbors locally. On receiving beacon packet, the node maintains the list of neighbors with their information (i.e., NeighborID, Position, SendToDelay, ExpireTime). Position field contains the distance from each neighbor to the destination node.

When a node receives an ACK form its neighbor node as a reply of transmitted data packet, it calculates the delay bound by calculating the elapsed time during transmission of data packet and reception of ACK. SPEED uses single hop delay that is the delay across one route.

The SNGF chooses the neighbor node, which achieves speed requirement according to the calculated delay values. If SNGF did not find the neighbor node based on delay values, then it checks the relay ratio of the node.

The relay ratio of the node is provided by Neighborhood Feedback Loop (NFL) module. The NFL calculates the relay ration by looking the miss ratios of the neighbors of a node and is transmitted to the SNGF module. The packet will be dropped, when relay ratio of the node is less than a randomly generated number (i.e., b/w 0 and 1).

Backpressure-rerouting module prevents voids, if a node fails to find next hop node. In SPEED, this module is used to remove congestion by transmitting messages back to source and use new routes to reroute messages.

Although SPEED protocol achieves the goal end-to-end delay bound data transmission, but it does not consider energy of sensors which affects network lifetime.

2.1 Improvements Considering Energy

The existing SPEED protocol considers Quality of Service (QoS) parameters for delay bound. In original SPEED protocol, the routes are always taken based on delay bound from source to destination. In the existing SPEED protocol, energy efficiency is not taken in account. However, in WSNs, the energy efficiency is an important issue. Therefore, we propose a modified SPEED protocol for WSNs. In our proposed modified SPEED protocol, routing decision will be taken based on two metrics: remaining energy of neighbor node to balance energy load on sensor nodes and end-to-end delay to maintain QoS requirements. This called as weight and is calculated by adding delay with remaining energy of neighbor node. The Figure 2 illustrates the use of weight notion in modified SPEED protocol. (A is source node and D is destination node)

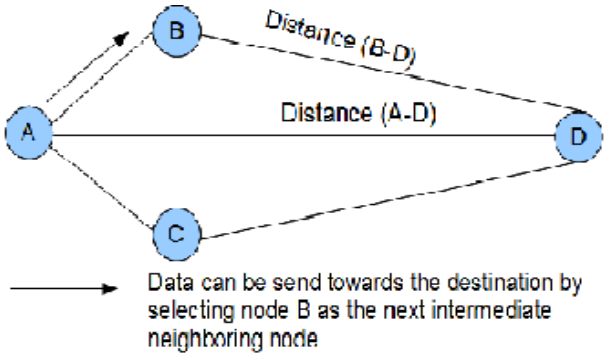


Figure 2: Modified SPEED protocol

Suppose, node A (source node) sends data packet to D (Destination node). **Node A will select its intermediate neighboring node according to estimated weight over link calculated in terms of energy strength of neighboring node. Following relationship was used to find the energy strength of each node.**

$$\textbf{weight} = \frac{e}{dv} \qquad \text{..... (2)}$$

where,

d = distance between intermediate neighboring node to destination node (m)

v = speed (m/s)

e = residual energy of intermediate neighboring node (J)

The sensor node with the greatest value (weight) will be next hop. As a result the sensor network lifetime is increased with greater throughput in QoS and energy load balancing way.

3 PERFORMANCE EVALUATIONS

3.1 Network Architecture

In the paper, network with 100 sensor nodes and single base station are deployed in area of 100 x 100 m^2. All sensor nodes are homogeneous.

3.2 Simulator

WSNet simulator [28] is used as a simulation platform to evaluate the proposed. WSNet is the modular event-driven wireless network simulator, developed at the CITI Laboratory of INSA Lyon. It is largely similar to other event-driven simulators such as ns2, JiST, GloMoSim, GTNetS, omnet++, though it differentiates itself with various functionalities, a precise radio medium simulation and the simulated node internals. Node, environment and radio medium blocks are developed in independent dynamic libraries. Moreover, the addition of new models does not require modifying the core of WSNet and can be done easily.

3.3 Simulation parameters

Table 1. Simulation parameters

Parameter	Value
Time for Simulation	50 to 500 seconds
Environment area (m x n)	100 x 100 m^2
Number of Nodes	100 nodes
Node deployment	Random
Performance parameters	Remaining Energy, Packet reached at the destination
Type of Antenna	Omni-directional
Medium Access Control Layer	802.11
Energy at each sensor node (initially)	1 J
Energy to transmit and receive a packet	0.003 Joules
Transmission range	30 m
Packet size	128 bytes

3.4 Simulation Results

In this subsection, we show the energy efficiency of modified SPEED protocol by comparing with the original SPEED protocol. For the evaluation, we consider remaining energy and the number of packets reached at the destination. We run simulations for 50 times and plotted the average results in the graphs.

In first scenario, we compared the remaining energy of sensor nodes by varying simulation time in our proposed modified SPEED scheme to original SPEED scheme. We determine total residual energy of nodes (i.e., remaining energy of nodes) and considered that as the metric to prove energy efficiency of our proposed protocol. Figure 3 shows the simulation results.

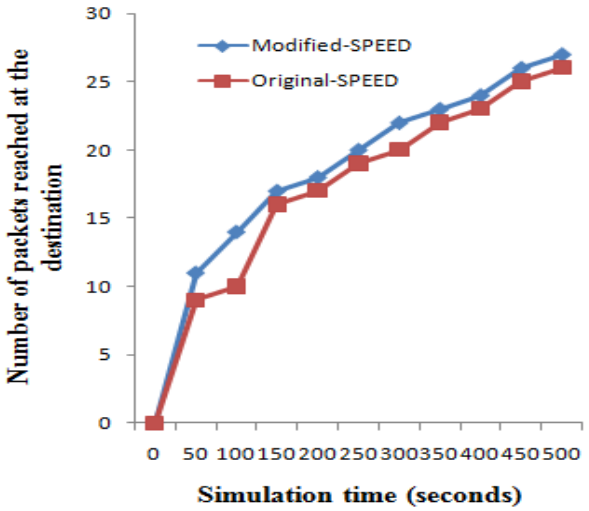


Figure 3: Total remaining energy

From Figure 3, it is proved that our proposed modified SPEED scheme saves more energy than the original SPEED scheme which will lead to an increase in the network lifetime.

In the second scenario we compared the packets reached at the destination by varying simulation time in our proposed modified SPEED scheme to original SPEED scheme. Figure 4 shows simulation results.

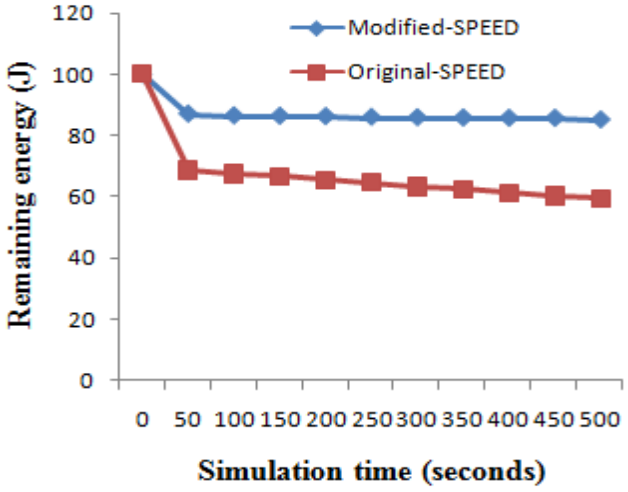


Figure 4: Packets reached at the destination

From Figure 4, it is proved that our proposed scheme is better than the original SPEED scheme.

4. CONCLUSIONS

In the paper, we have demonstrated most popular protocol for data routing in WSNs, called SPEED protocol that considers QoS parameters. Followed by an overview of SPEED protocol implementations, then we proposed a modified version of SPEED protocol. Through simulation results, we observed the performance of proposed protocol. Compared to the Original SPEED, 'Modified SPEED' conserves energy. Furthermore, the throughput (i.e., number of packets reached at the destination) of proposed protocol is more than the original protocol. That means the modified version of SPEED protocol outperforms the original version of SPEED protocol.

REFERENCES

[1] Shio Kumar Singh, M. P. Singh, D. K. Singh, "Applications, Classifications, and Selections of Energy-Efficient Routing Protocols for Wireless Sensor Networks", international Journal of Advanced Engineering Sciences and Technologies (IJAEST), Volume: 1, Issue No: 2, ISSN: 2230-7818, 2010.

[2] David W.Carman, Peter S. Kruus and Brian J.Matt, "Constraints and Approaches for distributed sensor networks security", Technical Report No.00-010, NAT Labs, Network associates Inc., Glenwood, MD, USA, 2000.

[3] M. Radi, B. Dezfouli, K. A. Bakar, and M. Lee, "Multipath Routing in Wireless Sensor Networks: Survey and Research Challenges", Sensors, Volume: 12, Number: 1, pages: 650-685, 2012.

[4] J. Kulik, W. Heinzelman and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks", in the Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mo- biCom99), 1999.

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: a Scalable and Robust Communication Paradigm for Sensor Networks", Proc. ACM Mobi-Com, Boston, MA, Pages: 56-67, 2000.

[6] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks", 1st Wksp. Sensor Networks and Apps, Atlanta, GA, 2002.

[7] C. Schurgers and M.B. Srivastava, "Energy-efficient routing in wireless sensor networks", Proc. Commun. for Network-Centric Ops, Creating the Info.Force, McLean, VA, 2001.

[8] Y. Yao and J. Gehrke, "The cougar approach to in network query processes in sensor networks", 2002.

[9] N. Sadagopan et al., "The ACQUIRE mechanism for efficient querying sensor networks", in the Proceedings of the First International Workshop on Sensor Network Protocol and Applications, Anchorage, Alaska , May 2003.

[10] R. Shah and J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks", in the Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), 2002.

[11] J. Kulik, W.R. Heinzelman, and H. Balakrishnan,"Negotiation-based protocols for disseminating information in wireless sensor networks", Wirel, Netw, Pages: 169-185, 2002.

[12] F. Ye, A. Chen, S. Lu, and L. Zhang, “A scalable solution to minimum cost forwarding in large sensor networks", In Proceedings of 10th International Conference on Computer Communications and Networks, Scottsdale, AZ, USA, Pages: 304-309, October 2001.

[13] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless sensor networks", in the Proceeding of the Hawaii International Conference System Sciences, Hawaii, January 2000.

[14] S. Lindsey and C. S. Raghavendra, “PEGASIS: Power Efficient GAthering in Sensor Information Systems", in the Proceedings of the IEEE Aerospace Conference, Big Sky, Montana, March 2002.

[15] A. Manjeshwar, D.P. Agarwal, “TEEN: A routing protocol for enhanced efficiency in wireless sensor networks", In Proceedings of 15th International Parallel and Distributed Processing Symposium, San Francisco, CA, USA, April 2001.

[16] A. Manjeshwar and D. P. Agrawal, “APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks", in the Proceedings of the 2nd International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile computing, April 2002.

[17] W. Ge, J. Zhang, G. Xue, \Joint clustering and optimal cooperative routing in wireless sensor networks", in Proceedings of International Conference on Communication, Beijing, China, May 2008.

[18] N. Dimokas, D. Katsaros, Y. Manolopoulos, “Energy-efficient distributed clustering in wireless sensor networks", Journal of Parallel and Distributed Computing, Volume: 70, Issue: 4, April, 2010.

[19] Trong Duc Le, Ngoc Duy Pham, Hyunseung Choo, “Towards a distributed clustering scheme based on spatial correlation in WSNs", in Proceedings of International Wireless Communications and Mobile Computing Conference, 2008.

[20] J.Chen, H. Shen, “MELEACH: an energy-efficient routing protocol for WSNs", Wireless Communications, Networking and Mobile Computing, 2008.

[21] H. Lu, J. Li, G. Wang, “A novel energy efficient routing algorithm for hierarchically clustered wireless sensor networks", in Proceedings of 4th International Conference on Frontier of Computer Science and Technology, December 2009.

[22] Y. Xu, J. Heidemann, and D. Estrin, “Geography-informed energy conservation for Ad-hoc routing", in the Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'01), Rome, Italy, July 2001.

[23] Y. Yu, D. Estrin, and R. Govindan, “Geographical and Energy-Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks", UCLA Computer Science Department Technical Report, May 2001.

[24] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, “SPAN: an Energy-efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", Wireless Networks, Volume: 8, Pages: 481-94, September 2002.

[25] F. Kuhn, R. Wattenhofer, and A. Zollinger, “Worst-case optimal and average-case efficient geometric ad-hoc routing", In Proceedings of 4th ACM International Conference on Mobile Computing and Networking, Annapolis, MD, USA, Pages: 267-278, June 2003.

[26] T. He et al, “SPEED: A stateless protocol for real-time communication in sensor networks”, in the Proceedings of International Conference on Distributed Computing Systems, Providence, RI, May 2003.

[27] E.Felemban, C.G. Lee, E. Ekici, “MMSPEED: Multipath Multi-SPEED Protocol for QoS Guarantee of Reliability and Timeliness in Wireless Sensor Networks”, IEEE Trans, Mobile Computing, volume: 5, pages: 738-754, 2006.

[28] WSNet simulator, available: http://wsnet.gforge.inria.fr/

ANALYSIS OF POWER OPTIMIZATION TECHNIQUES IN VISION BASED ENVIRONMENT MONITORING SYSTEMS

Umair Ali Khan*, Fareed Hussain Mangi**, Intesab Hussain Sadhayo*

ABSTRACT

Intelligent environment monitoring confronts the major challenge of electrical power minimization with the constraints of real-time response and high quality of service. Such systems are mostly battery-operated and typically perform the tasks of monitoring various (complex) physical phenomena (e.g., forest fire, solar radiation, crowed monitoring, intelligent traffic surveillance, border security, etc) in a hazardous environment and must have adequate processing capabilities to meet the real-time response requirements. With these capabilities being the focus of design, high power consumption is inevitable which is not affordable in the intended environments of operation. A system must be designed with power-efficient components as well as online power optimization strategies. This paper presents a critical analysis of the requirements and constraints for power and resource awareness in environment monitoring systems. We classify the existing power management techniques used in vision-based environment monitoring systems and identify the set of features that serves as a criterion to evaluate the performance of a technique. Our critical analysis and experiments with carefully designed systems demonstrate that a machine-learning based power optimization strategy gives optimal performance in dynamic environments and outperforms other techniques.

Keywords*: Power optimization; environment monitoring; visual sensor networks.*

1. INTRODUCTION

Due to the increased popularity and efficacy of visual sensors over the past few decades, intelligent environment monitoring of hazardous or rough environments is preferably performed using vision-based systems that are mostly battery operated. The systems developed in this context cover a wide range of applications which include monitoring indoor environmental state of large communication places, monitoring human behavior and privacy according to environmental perspective, detecting intrusions, object recognition and tracking, traffic monitoring, etc (to name a few). The advantages of a vision-based system are the following.

(i) Images contain a lot of information that can be processed and manipulated in a number of ways.
(ii) Heterogeneity of visual sensors can be used to perform a variety of tasks by a single system.
(iii) An environmental scene can be analyzed from different perspectives using image processing techniques.
(iv) The use of bulky and expensive visual sensors can be avoided.

However, managing power consumption of the individual visual sensor nodes in a vision-based system, primarily with the goal of extending their lifetime, is widely recognized as a fundamental challenge in intelligent environment monitoring. Since the monitoring systems are usually desired to be highly responsive (i.e., minimum response time at the detection of a particular event), a high processing capability is required, which in turn results in increased power consumption. Therefore, apart from high demands in computing performance, power-awareness is also of particular interest in vision-based intelligent environment monitoring systems.

The power optimization techniques proposed in this context usually fall into two categories: (i) static techniques, such as design-time synthesis and compilation for low-power consumption, and (ii) dynamic techniques that exploit runtime behavior of visual sensors to optimize power consumption [1][2]. These techniques are collectively known as Dynamic Power Management (DPM) that dynamically switches the power states of a system (or its individual components) based on the system's workload history, prediction of the future activity, processing requirements and desired performance.

Power management policies exploit the fact that an environment monitoring system is not always intended to deliver maximum performance, especially when the events occur at low rate or a high quality of service is not required. Therefore, these policies attempt to find the periods in which the system serves no or light workload. Based on this knowledge, some or all components of a system are switched to low-power states.

* Department of Computer Systems Engineering

** Department of Energy & Environment Engineering,

Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan.

A good DPM policy should follow a proactive approach, rather than relying only on a reactive strategy. Based on the current environmental conditions, the DPM policy should anticipate the future usage pattern of a system and should take proactive measures for power management. This extends the scope of DPM from statically optimized resources and power management to an intelligent and adaptive decision-making strategy.

The selection and/or formulation of a DPM policy for a vision-based system is not trivial due to the following challenges.

(i) Multiple power states of different components on a visual sensor node (e.g., processing, idle, sleep, off, etc) have different power consumption and performance levels. Therefore, an optimal tradeoff needs to be found.
(ii) The latency involved in switching among different power states is one of the major issues in a hard real-time environment. Furthermore, the power consumed to switch the power state of a component is also non-negligible. Therefore, it is very important to know when the state-switching latency and the resulting power consumption is effectively compensated and provide overall power savings.
(iii) An accurate prediction of workload results in a more effective state-switching policy, hence reducing the overall latency and increasing power savings [28][29]. Several studies show that the workload of even non-stationary environments exhibit long-term similarity [30][31]. However, accurate workload prediction in dynamic environments to capture the short-term variation is challenging.
(iv) Another challenge is to optimize the data transmission that consumes significant amount of power. When small response time and high quality of service are required, power optimization becomes challenging due to frequent transmissions.

This paper provides an extensive survey of the existing DPM policies in vision-based environment monitoring. We discuss the situations in which a specific DPM policy performs well and the factors that lead to the degradation of a DPM policy. Since real environments are highly dynamic, an important aspect of our investigation is to highlight the DPM policies that can deal with non-stationary request patterns. We identify several performance parameters that define the optimal behavior of a DPM policy. Using these parameters as evaluation criteria, we justify why static power management policies or those requiring a specific model of the system may not always lead to optimal behavior. This rigorous analytical survey and the deduced conclusions lead us to highlight the potential strategy for the DPM of a vision-based system running in a highly dynamic environment. We justify why the optimal behavior in a dynamic environment can be achieved only with an online machine-learning based DPM policy.

Figure 1 shows our vision-based traffic monitoring system based on heterogeneous visual sensors. The system performs online power optimization using machine-learning algorithms.

Figure 1: Our vision based traffic monitoring system under power management [32]

2. DPM REQUIREMENTS

Independent of the type of environment a system is intended to operate in, formulating and implementing a DPM problem for the system is not trivial. A DPM problem requires taking into consideration a number of factors that play critical roles in the success of a DPM policy. In general, a good DPM policy must consider the following factors.

2.1 Identifying the Culprit Components

In most cases, it is not necessary to target the whole system for power management. Instead, the power consumption can be optimized by targeting specific component(s) of the system. For example, in most of the mobile computing devices (e.g., cell phones), the display is the most power-consuming component [33]. A proper power management of the display component results in a significant power savings in the overall system. Therefore, before implementing a DPM policy, the most power-hungry elements in a system must be identified.

2.2 Environmental Uncertainty

The request processing pattern of a complex system is usually highly dynamic and dependent on the application, input data and the user context. The variations in the workload bring about changes in the system's usage pattern and substantially influence the power consumption and performance. Therefore, an important feature of a DPM policy is to learn the environment, adapt to different hardware systems, and operate under different working environments [34].

2.3 Power-Performance Tradeoff

Minimizing power consumption results in an inevitable performance degradation in every computing system.

Power optimization becomes a challenge for the systems where a real-time response is required. However, for the systems that are required to deliver a soft real-time response, performance can be traded for low power consumption. For such systems, a number of optimal solutions corresponding to different power-performance settings exist. A good DPM policy should allow the user to flexibly control the power-performance tradeoff to select one of these optimal solutions.

2.4 Transition Costs

Switching a device into low power or non-operational mode and then switching it back to an active (operational) mode results in a sizable and non-negligible amount of power consumption as well as latency. A DPM policy that tends to switch off a device too frequently during the periods of high activity, in fact results in extra power consumption and performance degradation. Therefore, a DPM policy should be able to find an appropriate balance among the state transitions. The decisions of switching the power states of a system should be taken only when the resulting power-consumption and latency can amortize the power savings achieved by the state transition.

2.5 Processing Requirements, Deadlines and Preferences

A DPM policy should be able to identify the processing requirements of individual tasks submitted to the system. Some requests may have stringent deadlines and should be processed with highest priority. In such cases, power consumption can be sacrificed to deliver a real-time response. Likewise, during different periods of the system's operation, the processing requirements and performance preferences may change. During the peak hours of a system's operation, the DPM policy should prefer high performance at the expense of higher power consumption.

3. DYNAMIC POWER MANAGEMENT TECHNIQUES

Several power management techniques have been studied for environment monitoring systems that can be broadly classified into two categories: adaptive (online) and non-adaptive (static). Essentially, all approaches determine how long a system should remain in an idle state before transiting to a low power (or off) state. Non-adaptive approaches statically determine the idle periods based on the properties of the system and do not consider the arrival rate of requests. Whereas, adaptive power management approaches compute the idle periods dynamically based on the history of incoming requests [35]. Adaptive power management approaches tend to perform better by making appropriate decisions over time based on the change in the load of the system [36]. Figure 2 provides a classification of existing power management techniques for environment monitoring systems.

Before analyzing the existing DPM policies individually, we prefer to present a generic DPM model shown in Figure 3. A particular objective to present this generic DPM model is to aggregate all the necessary components in a single (conceptual) model and then mapping the existing DPM approaches to this model. This helps us to highlight the salient features a particular DPM policy lacks or possesses. The conceptual model presented in Figure 3 depicts a system where the requests can be buffered before processing and an appropriate power-performance criterion can be selected specific to the environment and/or the system. The service requestor in Figure 3 is an image visual sensor that captures the events and buffers them in the service queue. The processing device, referred to as the service provider in this model, retrieves the requests from the service queue and processes them. The power manager is the DPM policy which switches the power modes of the service provider based on the observations received from all the components.

3.1 Greedy Policy

The simplest power management policy is to switch a device to a low-power state as soon as it is idle [37]. The greedy policy does not analyze the past workload of the system and the additional power and/or latency resulted by frequent state-transitions in case of higher workloads.

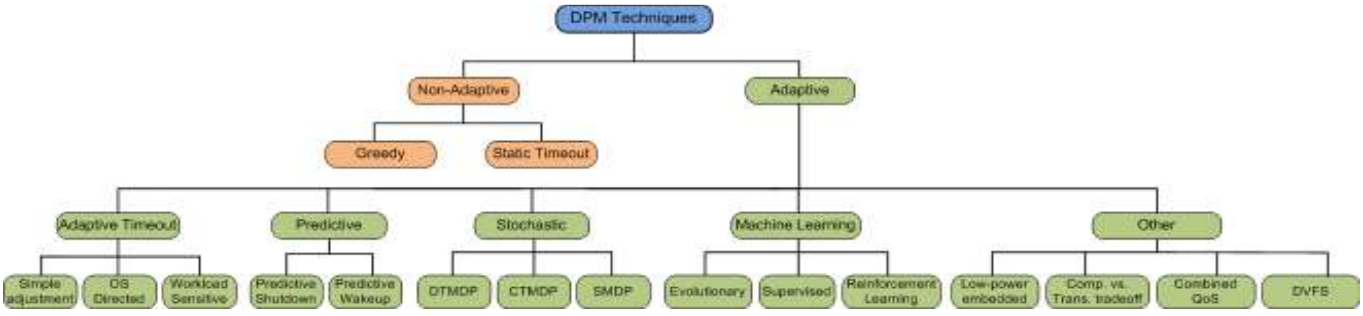


Figure 2: Classification of dynamic power management

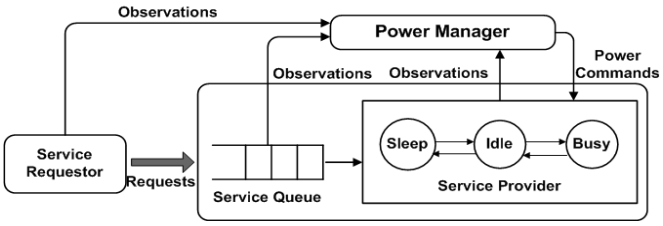


Figure 3: A generic model for power management

As long as the requests' inter-arrival times are long, the greedy policy performs well, as it can amortize the infrequent state-transitions cost in this case. However, real input patterns are usually non-stationary and it is not efficient to turn off a device during a bursty usage pattern. A DPM policy must have a mechanism to detect a period of inactivity to change the power mode of the device.

3.2 Timeout Policies

Instead of an immediate shutdown (or switching a device to low-power mode as soon as it is idle), the timeout policies assume that if a device is idle for a certain period of time, it would remain idle for the same period of time in future and it is relevant to switch it to low-power mode. The advantage of the timeout policy over the greedy policy is that the additional cost of power consumption and latency during the state transitions can be avoided when the workload is high. The two variants of timeout policies are static and adaptive policies [38]. A user defined fixed timeout is used in static policies. Due to their simplicity, static timeout policies are widely used in many commercial appliances, e.g., mobile phones, monitors and hard disks. An adaptive timeout scheme is more effective because it adjusts the timeout threshold according to the system's workload history.

Without incorporating an accurate workload prediction, timeout policies turn out to be too naïve. Once the systems becomes idle and starts a timer, it consumes power until the timeout expires. Furthermore, these policies are implementable only at the idle (active) state of a system and cannot deal with non-operational or sleep states. Additionally, timeout policies lack the mechanism for controlling the power-performance tradeoff.

Figure 4 shows the mapping of greedy and timeout policies to the generic DPM model shown in Figure 3.

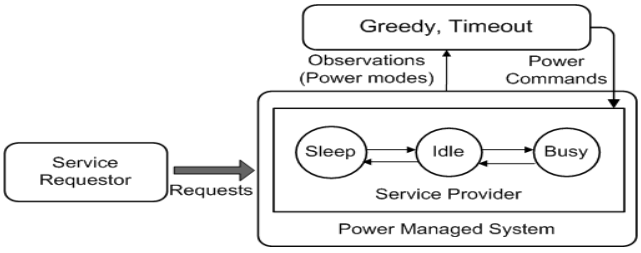


Figure 4: Mapping of greedy and timeout policies to the generic DPM model.

3.3 Predictive Policies

Predictive policies analyze the history of an environment monitoring system's workload and attempt to predict the future idle periods of the system using statistical approaches. The predictive techniques are extensively researched and used for a number of applications [2][5][6][10][11][12]. The basic approach of all the predictive policies is to use the correlation between the past usage pattern of a system and predict the arrival time of the next event. These policies predict the length of an idle period before it starts. If an idle period is predicted to be longer than a certain threshold (called break-even time), the device enters into a low power state right after it is idle. In order to predict the idle periods, several techniques have been introduced. One such technique uses adaptive learning trees [10] to encode the sequence of idle periods into tree nodes. This policy predicts the length of an idle period with finite-state machines. If an idle period is predicted to be longer than a given break-even period, the confidence level increases; otherwise, the confidence level decreases [2].

Some other techniques use mathematical predictors on the history of system's workloads in order to predict the future workload. The simplest of such predictor uses the weighted average workload filter that predicts the future idle period based on the average of past idle periods. A better variant of this technique is the moving average filter.

Exponential weighted average of past idle periods to predict the future period of inactivity is also studied. A better approach uses least mean square filter that works as an adaptive filter whose coefficients are modified based on the prediction error. In this sense, it works better than the exponential weighted averaging filter [5].

For the environment monitoring systems where the visual sensor nodes have a high degree of uncertainty in the observation, computation and communication, predictive policies do not perform impressively. Moreover, these policies cannot control the power-performance tradeoff. Figure 5 shows the mapping of predictive policies to the generic DPM model.

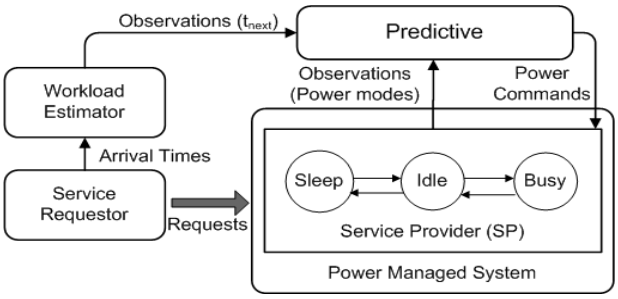


Figure 5: Mapping of predictive policies to the generic DPM model.

3.4 Stochastic Policies

Stochastic policies are based on the probabilistic model of a system's behavior with the associated uncertainty related to the system's inputs. This mathematical model is then used to formulate an optimization problem that is solved during the optimization process to obtain an optimal DPM strategy [3]. The modeling of power states of a device, transition mechanism, and its service queue is generally performed using Markov Decision Process (MDP). An MDP represents a system in which the next state is independent of the history of state transitions and depends only on the current state [9]. Considering the uncertainty in the environment, communication, computation load, occurrence of events, and nature of events in environment monitoring systems, several variants of the stochastic approaches can be used to describe the system and find the optimal DPM policy. These variants include discrete-time, continuous-time and semi-Markov decision processes (represented by DTMDP, CTMDP and SMDP, respectively in Figure 2).

A simple 2-state MDP is shown in Figure 6 where the system has two power states, i.e., active and sleep. The arrows represent the state transitions. Every transition is assigned a certain probability with which the power manager decides to change the state of the system or to remain in the same state.

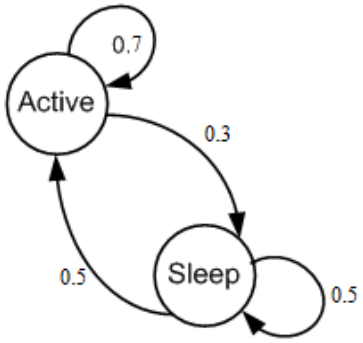


Figure 6: A 2-state Markov decision process

Stochastic techniques support the request queuing as described in the generic model presented in Figure 3 and hence are able to trade power with performance. However, stochastic techniques require the mathematical formulation of the system's model (as shown in Figure 6) which is not only nontrivial, but also makes it environment specific.

3.5 Low-Power Embedded Designs

Low-power embedded designs are also of particular interest in environment monitoring systems. These systems comprise low- or medium resolution image visual sensors (mostly CMOS) integrated with an embedded computing platform [13][14][15][16][17]. The key power management strategy in such systems is based on system-level dynamic power management because embedded systems provide good power management in terms of better control and access to various system components like system buses, memory, communication units and power modes of processor and other components. The power management policy in such systems is usually implemented at operating system level that deals with changing the power modes of individual system components based on the occurrence of events, available resources and/or processing requirements. Since the vision-based environment monitoring usually requires high-performance on-board image and video processing, these systems also use dedicated FPGAs or DSPs to perform image acquisition, compression, encoding and decoding operations. A generic representation of such systems is shown in Figure 7.

Unlike the additional DPM strategies implemented at application level, the DPM techniques applied at the OS level usually do not deliver aggressive power savings. Although such systems usually prove to be highly responsive, power savings are not significant.

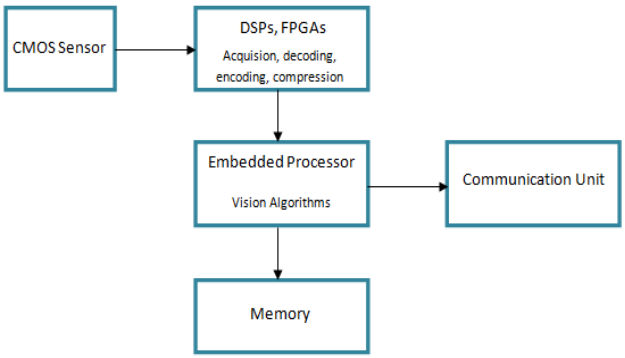


Figure 7: Generic representation of surveillance systems based on embedded design

3.6 Computational load and transmission power trade-off

Data communication in vision-based environment monitoring systems is more power consuming than processing and sensing. Therefore, significant power savings can be obtained by minimizing the frequency of transmission [18][19]. Some approaches focus on finding an appropriate tradeoff between processing and communication. In order to reduce the transmission load, one of the solutions is to perform the local data/image compression. However, the power required for image compression is comparable to the one consumed in transmission. In [18], the compression rates, processing times and power consumption of different compression algorithms are studied. However, the performance of these algorithms heavily depends on the application and the overall visual sensor configuration. This approach can

be implemented just as a part of the overall dynamic power management strategy and requires a detailed investigation of the power consumptions and processing load of different compression algorithms for a particular setup.

3.7 Quality of Service (QoS) Based DPM

This technique combines power optimization with QoS adaptation [20][21]. The degree of freedom in adapting QoS parameters strictly depends on its designated application. If it is acceptable for the user and non-critical in the context of the application, the level of QoS may be manipulated and power can be saved. Thus, this method is a tradeoff between the QoS and power consumption.

This technique dynamically changes the power modes of individual components (e.g., visual sensors, processing units, network devices) based on the required QoS level. The system runs in normal mode when no event is detected and hence delivers minimum QoS level with maximum power savings. At the detection of an event, adequate QoS parameters are selected for each component. However, this approach does not take into account the latency involved in switching the power modes of different components. Since this approach is aimed to implement on a hard real-time system, this latency cannot be ignored.

3.8 Dynamic Voltage and Frequency Scaling (DVFS)

Another wide area of research in dynamic power management is dynamic voltage and frequency scaling. DVFS exploits the fact that the peak clock frequency of a processor is proportional to the supply voltage, while the amount of dynamic power required for a given workload is proportional to the square of the supply voltage [22]. DVFS varies the clock frequency of the processor and its supply voltage in such a way that the requests are processed within their deadlines or with minimum latency. This requires a proper scheduling of processor's clock frequency and supplied voltage. The task of scheduling processor's frequency and voltage is performed by the operating system that has a global view of the resource usage, requirements and workload. A number of techniques have been proposed for predicting the processor's operating frequency and voltage at the operating system level [23][24][25][26][27]. The rationale in all these techniques is to analyze the processor's workload history and to predict the future workload in order to set an appropriate frequency and voltage. However, DVFS cannot control the power-performance tradeoff.

3.9 Machine-learning Based Policies

Machine-learning based techniques are of particular interest to find patterns of a system's usage and its workload. Predicting a system's periods of inactivity is also performed using genetic (evolutionary) algorithms [39]. This approach exploits the relationship between adjacent periods of activity and inactivity to predict the future period of inactivity and decides if it is appropriate to switch the system to low-power state. Nevertheless, this approach works no better than the previously described predictive policies.

Some supervised learning techniques for power management [40][41] use offline data analysis to extract essential features of the system which collectively describe the power profile of the system. The system is trained with these features to generate a classifier (e.g., Bayesian) to predict the system's performance state from new observations and then using this predicted state to look up the optimal power management action from a pre-computed policy table. The offline data collection, analysis, and identification of the essential features is cumbersome and requires the same effort as in the case of stochastic policies.

A variant of unsupervised learning, called Reinforcement Learning (RL), is found to be significantly more effective to learn a system's environment and take better actions for power optimization. Power optimization with RL does not require a priori model of the system. Instead, the main theme is to learn the environment of a system by starting with no or little knowledge and trying out various alternatives in each state of the system until good actions for each state are learnt.

RL is originally applied to MDPs and an power optimization problem does not essentially represent an MDP. However, some model-free, RL based DPM approaches proposed in [42][43] prove that RL can be successfully applied to DPM problems. RL is first implemented for DPM in [42], where some pre-selected policies, referred to as experts, are selected under different operating conditions for power optimization. Each expert is assigned a weight that varies throughout the online learning process and represents the effectiveness of an expert in a given set of conditions. In each state of the system, an expert with the highest weight is selected with the highest probability. This approach does provide an optimal power management technique, but its performance is strongly dependent to the selected experts.

Another model-free, RL-based DPM technique is proposed in [43] that does not require a set of experts and is able to effectively control the power-performance tradeoff. Nevertheless, this technique works with a huge state-space and is based on a discrete-time stochastic process. Hence, it is not memory- and computing-efficient.

The authors in [44] propose an improved, continuous-time RL based policy by incorporating a workload

prediction.

A naïve Bayes classifier predicts the inter-arrival times of future events by the history of past inter-arrival times. However, the implementation of this learning policy is done on a wireless LAN card that sends/receives data from the protocols generating regular traffic having higher degree of predictability. This workload prediction cannot deliver acceptable accuracy in non-stationary environments.

We address this issue in [45][46] using the DPM model shown in Figure 3. We specifically target the non-stationary workloads with a machine-learning approach that does not require any prior model. We use a multi-layer neural network with back-propagation algorithm to provide estimated workload information to the learning algorithm. With the predicted workload information, this approach selects appropriate timeout policies in the idle state of a system. The wakeup decision in sleep state is made on the basis of service queue populated with certain number of requests. In contrast to the existing DPM approaches that focus on an immediate response from the service provider as soon as there is a request in service queue, we opt to postpone the request processing to explore a deeper power-performance tradeoff. The time spent by the system in active and inactive states is controlled by an adjustable power-performance tradeoff parameter. Workload prediction using a multi-layer neural network achieves higher accuracy with the non-stationary data and the algorithm is capable of exploring the tradeoff in the power-performance design space. However, this approach does not ascertain the upper and lower bounds of average latency in request processing and also results in high latency when the workload is low.

We further address this issue in [28][29] and propose an RL based policy called Online Learning of Timeout Policies (OLTP) for the power management of environment monitoring systems. This policy targets both the idle and sleeps states to give a better power-performance tradeoff. Additionally, we extend the scope of OLTP by enabling it to handle multiple sleep and idle states of the modern computing systems. Our experimental results show that a machine-learning based policy is able to learn the optimal timeout values for each state of the system based on the environmental changes and system's workload. Our rigorous comparative analysis also shows that the OLTP outperforms existing DPM policies for environment monitoring.

Figure 8 shows the mapping of stochastic and machine-learning policies to the generic DPM model. Through our rigorous analysis, we identify a set of distinct features that serves as a criterion to evaluate the performance of a given DPM policy. The comparative analysis of the existing DPM policies for environment monitoring with respect to these features is given in Table 1. A YES or NO value followed by a (+) sign represents the merit of a DPM policy for the corresponding feature. It is evident that the machine-learning based DPM policy outperforms other policies.

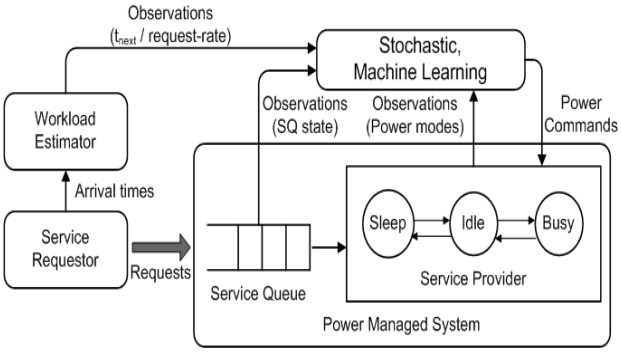


Figure 8: Mapping of stochastic and machine-learning policies to the generic DPM model.

4. CONCLUSION

In this paper, we performed a critical analysis of the power management techniques for vision-based environment monitoring systems. We described some essential environmental and system-level factors an power management policy should be focused on. Subsequently, we provided a detailed analysis of the existing power management policies and argued why a RL-based DPM approach can outperform other policies. An important part of our discussion involved the investigation of the applicability of the existing DPM policies to the generic DPM model described in Section 3. Furthermore, we highlighted essential features of a good DPM policy and compared the existing DPM policies with respect to the identified performance parameters.

REFERENCES

[1] Arnold Maier, "Dynamic power aware camera configuration in distributed embedded surveillance clusters", PhD Dissertation, Technical University Graz.

[2] Y. Lu, G. Micheli, "Comparing System-level power management policies", IEEE conference on Design & Test of Computers, 2001

[3] A. Fallahi, E. Hossain, "QoS provisioning in wireless visual sensor networks", IEEE journal of wireless communication, vol.14, No.1, pp.40-49, 2007.

[4] B. Brock, K. Rajamani, "Dynamic power management for embedded systems", IEEE SOC Conference, pp. 1-25, 2003.

[5] A. Sinha, A. Chandrakasan, "Dynamic power management in wireless visual sensor networks", IEEE Design & Test of Computers, vol. 18, No. pp. 62-74, 2001.

[6] L. Benini, A. Bogliolo, G. D. Micheli, "A survey of design techniques for system-level dynamic power management", IEEE Transactions on very large scale integration (VLSI) systems, vol.8, No.3, 2000.

[7] G. Norman, D. Parker, M. Kwiatkowska, "Using probabilistic model checking for dynamic power management", Formal aspects of computing, vol. 17, No. 2, pp. 160-176, Springer-Verlag, 2005.

[8] Q. Qiu, Q. Wu, M. Pedram, "Stochastic modeling of a power-managed system – construction & optimization", IEEE Transactions on computer-aided design of integrated circuits & systems, vol.20, No.10, pp. 1200-1217, 2001.

[9] Guo, Xianping, and Onésimo Hernández-Lerma, "Continuous-time Markov decision processes", Springer Berlin Heidelberg, 2009.

[10] E. Y. Chung, L. Benini, G. D. Micheli, "Dynamic power management using adaptive learning tree", Proceedings of the 1999 international conference on computer aided design, pp. 274-279, 1999.

[11] T. Simunic, L. Benini, P. Glynn, G. Micheli, "Dynamic power management for portable systems", Proceedings of the 6th annual international conference on mobile computing and networking, MobiCom, pp. 11-19, 2000.

[12] C. Hwang, A.C. Wu, "A predictive system shutdown method for power saving of event-driven computation", ACM Journal on design automation of electronic systems, vol.5, No. 2, pp. 226-241, 2000.

[13] Z. Y. Cao, Y. Ji, M. Z. Hu, "An image Visual sensor node for wireless visual sensor networks", Proceedings of the International Conference on Information Technology: Coding and Computing", pp. 740-745, 2005.

[14] M. Bramberger, R. P. Pflugfelder, B. Rinner, "A smart camera for traffic surveillance", Proceedings of the first workshop on Intelligent Solutions in Embedded Systems", pp. 153-164, 2003.

[15] C. B. Margi, R. Manduchi, K. Obraczka, "Power consumption tradeoffs in visual sensor networks", 12th Brazillian Symposium on Computer Networks, pp. 1-4, 2006.

[16] B. Brock, K. Rajamani, "Dynamic power management for embedded systems", IEEE International SoC conference, pp. 1-25, 2003.

[17] W. Feng, E. Kaiser, M. Baillif, "Panoptes: scalable low-power video visual sensor networking technologies", ACM Transactions on multimedia computing, communications and applications, vol.1, No.2, pp. 151-167, 2005.

[18] L. Ferrigno, S. Marano, A. Pietrosanto, "Balancing computational and transmission power consumption in wireless image visual sensor networks", IEEE Conference on virtual environments, human-computer interfaces, and measurement systems, pp. 1-6, 2005.

[19] V. Raghunathan, S. Park, M. B. Srivastava, "Power-aware wireless micro-visual sensor networks", IEEE Signal Processing Magazine, vol. 19, No. 2, pp. 40-50, 2002.

[20] A. Maier, B. Rinner, T. Trathnigg, "Combined dynamic power and QoS management in embedded video surveillance systems", Proceedings of the second workshop on intelligent solutions in embedded systems, pp. 63-77, 2004.

[21] A. Maier, "Dynamic power aware camera configuration in distributed embedded surveillance clusters", PhD Dissertation, Technical University Graz, 2006.

[22] Mudge. T, Ann Arbor, "Power: a first-class architectural design constraint", Computer, Vol. 34, No. 4, pp. 52-58, 2001.

Table 1: Comparative analysis of power management techniques in environment monitoring systems

	Performance constraint	Performance tradeoff	Workload handling	Queuing	Model-free	Non-stationarity	Offline analysis	Complexity
Static timeout	No (-)	No (-)	No (-)	No (-)	Yes (+)	No (-)	No (+)	Low
Adaptive timeout	No (-)	No (-)	Yes (+)	No (-)	Yes (+)	No (-)	No (+)	Low
Predictive shutdown	No (-)	No (-)	Yes (+)	No (-)	Yes (+)	Yes (+)	No (+)	Low
Predictive wakeup	Yes (+)	No (-)	Yes (+)	No (-)	Yes (+)	Yes (+)	No (+)	Low
Stochastic	Yes (+)	Yes (+)	Yes (+)	Yes (+)	No (-)	Yes (+)	Yes (-)	High
Low-power embedded design	N/A	No (-)	No (-)	No (-)	N/A	N/A	N/A	low
Computing-Transmission tradeoff	No (-)	No (-)	No (-)	N/A	N/A	N/A	N/A	high
QoS tradeoff	Yes (+)	Yes (+)	No (-)	No (-)	N/A	No (-)	N/A	high
DVFS	Yes (+)	No (-)	Yes (+)	No (-)	Yes (+)	No (-)	No (+)	low
Evolutionary	No (-)	No (-)	Yes (+)	No (-)	Yes (+)	No (-)	No (+)	Low
Supervised learning	Yes (+)	Yes (+)	Yes (+)	No (-)	Yes (+)	Yes (+)	Yes (-)	Moderate
Online RL	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	No (+)	Moderate

[23] D. Grunwald, P. Levis, K. Farkas, "Policies for Dynamic Clock Scheduling", Proceedings of the 4th conference on symposium on operating system design & implementation, pp. 1-6, 2000.

[24] K. Flautner, T. Mudge, "Vertigo: Automatic performance-settings for Linux", Proceedings of the 5th symposium on operating systems design and implementation, pp. 47-52, 2002.

[25] T. Pering, T. Burd, R. Brodersen, "The simulation and evaluation of dynamic voltage scaling algorithms", Proceedings of the international symposium on low power electronics and design, pp. 76-81, 1998.

[26] K. Govil, E. Chan, H. Wasserman, "Comparing algorithms for dynamic speed-setting of a low-power CPU", Proceedings of the 1st annual international conference on mobile computing and networking, pp. 13-25, 1995.

[27] M. Magno, A. Lanza, "Power aware multimodal embedded video surveillance", IEEE conference on VLSI system on chip, pp. 264-279, 2010.

[28] Khan, Umair Ali, and Bernhard Rinner, "Online learning of timeout policies for dynamic power management." ACM Transactions on Embedded Computing Systems, vol. 13, No. 4, pp. 1-25, 2014.

[29] U. A. Khan, F. A. Jokhio, I. H. Sadhayo, "Reinforcement Learning for dynamic power management of embedded visual sensor nodes", Mehran University Research Journal of Engineering & Technology, vol. 33, No.2, 2014.

[30] P. Bogdan and R. Marculescu, "Cyber physical systems: Workload modeling and design optimization", IEEE Transactions on Design & Test, vol. 28, No. 4, pp. 78–87, 2011.

[31] P. Bogdan and R. Marculescu, "Non-stationary traffic analysis and its implications on multi-core platform design", IEEE Transactions on Computer Aided Design of Integrated Circuits Systems, vol. 30, No. 4, pp. 508–519, 2011.

[32] U. A. Khan, B. Rinner, "A reinforcement learning framework for dynamic power management of a portable, multi-camera traffic monitoring system", In Proceedings of IEEE conference on Green Computing & Communication (GreenCom), pp. 557-564, Besancon, France, 2012.

[33] J. Lorch and A. Smith, "Software strategies for portable computer power management", IEEE Transactions on Personal Communications, vol. 5, No. 3, pp. 60–73, 1998.

[34] W. Liu, Y. Tan, and Q. Qiu, "Enhanced Q-learning algorithm for dynamic power management with performance constraint", In Proceedings of the Conference on Design, Automation and Test in Europe, pp. 602–605, 2010.

[35] D. Ramanathan, S. Irani, and R. Gupta, " An analysis of system level power management algorithms and their effects on latency", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 21, No. 3, pp. 291–305, 2002.

[36] M. Pedram, "Power aware design methodologies", Springer, 2002.

[37] M. Pedram, "Power optimization and management in embedded systems", In Proceedings of the ASP-DAC Design Automation Conference, pp. 239–244, 2001.

[38] Y. Lu and G. De Micheli, "Comparing system-level power management policies", IEEE Transactions on Design and Test of Computers, vol. 18, No. 2, pp. 10–19, 2001.

[39] K. Fei, T. Pin, Q. Shi, and L. Xiao, "Genetic algorithm based idle length prediction scheme for dynamic power management", In Proceedings of IMACS Multi-conference on Computational Engineering in Systems Applications, pages 1437 –1443, 2006.

[40] S. Mannor et al., "Machine-learning for adaptive power management", Autonomic Computing, vol. 10, No. 4, pp. 299–312, 2006.

[41] J. Hwisung and M. Pedram, "Supervised learning based power management for multi-core processors", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 29, No. 9, pp. 1395 –1408, 2010.

[42] G. Dhiman and T. Rosing, "Dynamic power management using machine-learning", In Proceedings of IEEE/ACM International Conference on Computer-Aided Design, pp. 747–754, 2006.

[43] Y. Tan, W. Liu, and Q. Qiu, "Adaptive power management using reinforcement learning", In Proceedings of International Conference on Computer-Aided Design, pp. 461–467, 2009.

[44] Y. Wang, Q. Xie, A. Ammari, and M. Pedram, "Deriving a near-optimal power management policy using model-free reinforcement learning and bayesian classification", In Proceedings of Design Automation Conference, pp. 41–46, 2011.

[45] U. Khan and B. Rinner, "Dynamic power management for portable, multi-camera traffic monitoring", In Proceedings of the 18th IEEE Real-Time and Embedded Technology and Applications Symposium (WiP), pp. 37–40, 2012.

[46] U. Khan, M. Godec, M. Quaritsch, M. Hennecke, H. Bischof, and B. Rinner, "Mobitrick: Mobile traffic checker", In proceedings of ITS World Congress, pp. 1-10, 2012.

RESOURCE ALLOCATION BASED ON RATES, LOAD AND THROUGHPUT FOR VIDEO TRANSCODING IN A CLOUD

Fareed Jokhio*, Umair Ali Khan*, Intesab Hussain Sadhayo*, Anees Ahmed Soomro*, Azhar Uddin Jokhio**

ABSTRACT

This paper presents resource allocation algorithms for video transcoding service on a cloud. The main objective of the proposed algorithms is to allocate and de-allocate Virtual Machines (VMs) horizontally in a cluster of video transcoding servers. For cost-efficiency and better utilization of resources, video segmentation at *group of pictures* level is used. With video segmentation, a video is split into smaller segments that can be sent for transcoding on any transcoding server. To demonstrate the efficiency of the proposed algorithms, a discrete-event simulation is used. The proposed algorithms are also compared with the existing video transcoding algorithms in a cloud computing environment. The existing VM allocation algorithms are based on accumulative play rate and transcoding rate, while our improved proactive VM allocation algorithms also take into account the overall computation load and system throughput. The results indicate that the proposed algorithms are more cost-efficient than the existing algorithms.

Keywords: *Video transcoding; cloud computing; resource allocation; load prediction*

1. INTRODUCTION

The cloud computing provides computing and storage services over the Internet. These services may include computer processing power, data storage space, software applications and networks. The provisioning of resources while needed and releasing the resources while not required is possible in a cloud computing environment. There exist a number of cloud computing service models such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Data as a Service(DaaS) [15].

In IaaS model, the service provider is responsible for the services including hosting and running of the applications, and maintenance of the equipment. In PaaS model, the service provider provides computational resources through an operating system. PaaS provides an environment to build and deploy applications. In SaaS model, the service provider provides implementation of a specific software application. The customer does not have to worry about getting a licence or getting acquinted with any other technical details. In DaaS model, the data is provided to the users on demand regardless of the physical location. The computing resources offered by IaaS clouds, such as computing units or virtual machines (VMs) [14], can be used to perform very intensive computations such as video transcoding [16] of several video streams simultaneously.

Video trasnscoding is a process in which a compressed video is converted from one format into another. It is needed because a video stored at the server side in a certain format might not be compatible with a device at the client side. A video usually requires a huge disk space. Therefore, it is always stored in compressed format such as MPEG-4 [17] and H.264 [18]. A video transcoding operation may change video format, frame resolution, frame rate, bit rate, or any combination of these.

The video transcoding requires intensive computation and it is not always possible to perform it on a client device such as a mobile phone in a real-time. The transcoding can be performed at server side or in a cloud computing environment and the transcoded video can be stored to avoid repeated transcoding operations. Cloud computing is getting popular due to its computing environments, software services and it can provide a VM with theoretically infinite capabilities. Therefore, it is suitable for video transcoding [3]. However, the number of video transcoding requests may vary with time. Hence, determining the number of resources required to perform computation in a cloud computing environment is still an open research problem.

Real-time transcoding of video streams requires a number of video frames according to the play rate of the stream to avoid violation of the client-side quality of service (QoS) requirements [6]. It is often possible to segment a video into small parts and perform the transcoding operation [12]. It helps to have better utilization of computing resources.

In this paper, we present resource allocation algorithms for video transcoding in a cloud computing environment. Load prediction with a load tracker and a load predictor [2] using linear regression model is used to predict the

* Department of Computer Systems Engineering, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan. Email: {fajokhio, umair.khan, intesab, anees.soomro}@quest.edu.pk

** Department of Electronics Engineering, MUET Jamshoro, Pakistan.

work load. Video segmentation is also used to share the computing resources. The proposed approach is evaluated in different experiments using discrete event simulation. The experiments show that the proposed algorithms are more cost-efficient as compared to the existing resource allocation algorithms.

We proceed as follows. In Section 2, we present the system architecture. Section 3 describes dynamic resource allocation algorithms. Section 4 presents our simulation results and Section 5 presents conclusion.

2. SYSTEM ARCHITECTURE

The system architecture of video transcoding in a cloud is shown in Figure 1. Users send requests to the streaming server which in turn sends them responses. Once the streaming server receives a request, it first checks weather the requested video is available in the video repository in the required format. If it finds the video in the repository, it sends it to the user. However, if the video is not available in the required format, it is sent for transcoding. As a video is sent for transcoding, it is segmented into small segments by the video splitter. A video consists of different types of frames having inter-dependencies among them. Therefore, segmentation is possible at key frames or *I* frames only. Other types of frames which require reference frames for decoding are *P* frames and *B* frames. The key frames are independent and do not require any other reference frame for decoding. Therefore, a video consists of *I* frames after a certain interval of frames. This is termed as a *Group Of Pictures (GOP)*. Distributed video transcoding with video segmentation at GOP level is used in [12][9]. These papers provide the implementation of a video transcoder using the Message Passing Interface (MPI) model and analyze various possible segmentation methods. In addition to the above works, several other works on video transcoding in the cloud computing [5][10][11] have adopted the video segmentation at GOP level. In our proposed approach, the video segmentation is also performed at GOP level. Once a video is segmented, it is sent to the *load balancer* for load balancing and scheduling.

The *load balancer* uses the shortest queue length policy for scheduling. The actual transcoding is performed at *transcoding server* and then the video segments are sent to *video merger* for merging. Finally, the streaming server sends the response to the users. This system architecture is similar to the techniques proposed in [5][10][11].

The resource allocation algorithms which are discussed in section 3 are implemented by the *master controller*. Based on the decisions of the *master controller*, the *cloud provisioner* provisions and terminates the VMs. There also exists a *load predictor* which performs the load prediction. It uses a two-step approach with a load tracker and a load predictor which was proposed by Andreolini and Casolari [2] to predict future load behavior under real-time constraints.

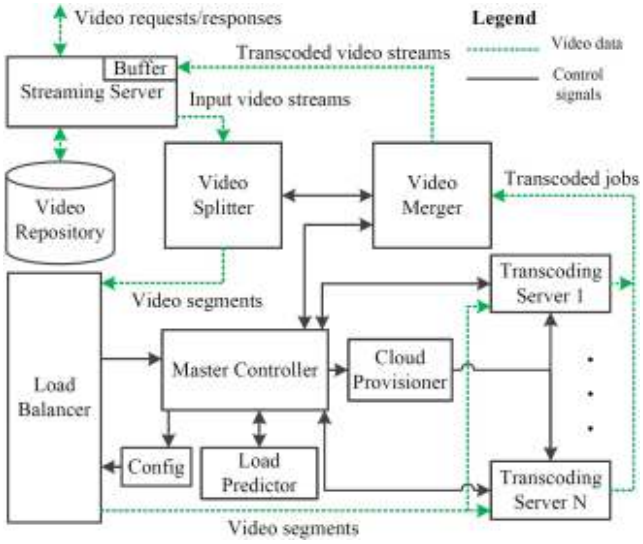


Figure 1: The system architecture

3. PROPOSED VM ALLOCATION AND DE-ALLOCATION TECHNIQUES

In this section, the proposed VM allocation and de-allocation algorithms are discussed which are based on accumulative play rate, accumulative transcoding rate, computation load and system throughput.

In [11], similar algorithms are proposed based on accumulative play rate and transcoding rate. However, resource allocation based on accumulative play rate and transcoding rate may have high provisioning cost. Additionally, it does not take into account the current computation load and the system throughput. The required transcoding rate could be lower due to improper load balancing. Therefore, provisioning more number of servers will result in a higher cost. In contrast, our proposed algorithms are based on accumulative play rate, transcoding rate, computation load and system throughput. The concepts and notations used in the algorithms are given in Table 1.

We apply a two-step load prediction approach [1][4][11] to predict the future workload. Our proposed technique tracks the current and past workload of the system and then predicts the future workload. As in [11], the system maintains a fixed number of VMs, termed as base capacity N_B. Therefore, our approach also maintains a fixed minimum number of servers.

The master controller collects the play rate $PR(t)$ and transcoding rate $TR(t)$ after some time interval. It performs the prediction and also calculates the predicted

transcoding rate $\hat{TR}(t)$. Since the provisioning of a VM requires some time [4], the algorithms avoid oscillations in the number of VMs [19] by delaying the new VMs allocation until the previous VM allocation operation is performed [8].

Notation	Description
$avgQJ(t)$	average queue length of all servers at t
$count_o(t)$	over allocation count at t
$N_P(t)$	number of servers to provision at t
$N_{P_Q}(t)$	number of servers to provision at t based on Queue length
$N_T(t)$	number of servers to terminate at t
$PR(t)$	sum of target play rates of all streams at t
$S(t)$	set of transcoding servers at t
$S_p(t)$	set of newly provisioned servers at t
$S_c(t)$	servers close to finish renting period at t
req_i	video transcoding $request_i$
C_j	$Class_j$ of video transcoding requests
$NSC_k(t)$	number of servers for $Class_k$ stream at time t
$NStrmC_k(t)$	number of streams in $Class_k$ at time t
$PRC_k(t)$	play rate of streams in $Class_k$ at time t
$avgTRC_k(t)$	average transcoding rate of $Class_k$ at time t
$NAS(t)$	total number of servers at time t
$S_t(t)$	servers selected for termination at t
$TR(t)$	total transcoding rate of all servers at t
$\hat{TR}(t)$	predicted transcoding rate of all servers at t
$RT(s,t)$	remaining renting time of server s at t
B_L	buffer size lower threshold in megabytes
$B_S(t)$	size of the output video buffer in megabytes
B_U	buffer size upper threshold in megabytes
$jobCD$	job completion delay
MQL_{UT}	maximum queue length upper threshold
N_B	number of servers to use as base capacity
RT_L	remaining time lower threshold
RT_U	remaining time upper threshold
$startUp$	server startup delay
$calcN_P()$	calculate the value of $N_P(t)$
$calcN_T()$	calculate the value of $N_T(t)$
$calcQN_P()$	calculate the value of $N_{P_Q}(t)$
$calRT(s,t)$	calculate the value of $RT(s,t)$
$calcNAS()$	calculate the value of $NAS(t)$
$delay(d)$	delay for duration d
$getPR()$	get $PR(t)$ from video merger
$getTR(s)$	get transcoding rate of server s
$get\hat{T}R()$	get $\hat{TR}(t)$ from load predictor
$prov(n)$	provision n servers
$select(n)$	Select n servers for termination
$sort(S)$	sort servers S on remaining time
$term(S)$	terminate servers S

Table 1: Summary of concepts and their notation for VM allocation algorithms

3.1 VM Allocation Algorithm

The VM allocation mechanism is described as Algorithm 1. In the algorithm, the play rate $PR(t)$ and total transcoding rate $TR(t)$ is computed. Then the predicted transcoding rate $\hat{TR}(t)$ of all streams is computed by the load predictor. It further checks the buffer size $B_S(t)$ where the transcoded video segments are stored.

Algorithm 1 VM allocation algorithm

1: **While true do**
2: $N_P(t) := 0, N_{P_Q}(t) := 0$
3: $PR(t) := getPR()$
4: $TR(t) := 0$
5:
6: $NAS(t) := calcNAS()$
for $s \in S(t)$ **do**
7: $TR(t) := TR(t) + getTR(s)$
8: **end for**
9: $\hat{TR}(t) := get\hat{TR}(TR(t))$
10: **if** $\hat{TR}(t) < PR(t) \wedge B_S(t) < B_L$ **then**
11: $N_P(t) := calcN_P()$
12: **end if**
13: **if** $avgQJ(t) > MQL_{UT}$ **then**
14: $N_{P_Q}(t) := calcQN_P()$
15: **end if**
16: $N_P(t) := N_P(t) + N_{P_Q}(t)$
17: **if** $len(S(t)) \cup N_P(t) > NAS(t)$ **then**
18: $N_P(t) := 0$
19: **end if**
20: **if** $N_P(t) \geq 1$ **then**
21: $S_p(t) := prov(N_P(t))$
22: $S(t) := S(t) \cup S_p(t)$
23: $delay(startUp)$
24: **end if**
25: **end while**

If the buffer size drops below the lower threshold B_L, the algorithms compute the number of VMs to provision $N_P(t)$ by using the following equation.

$$N_P(t) = \left\lceil \frac{PR(t) - \hat{TR}(t)}{\frac{TR(t)}{|S(t)|}} \right\rceil \quad (1)$$

Here $|S(t)|$ is the number of transcoding servers at time t. In addition to the buffer underflow, the resource allocation algorithm takes into account the average queue length of all transcoding servers $avgQJ(t)$. If it exceeds the maximum upper threshold MQL_{UT}, then, a number of servers is provisioned $N_{P_Q}(t)$ by using the following equation.

$$N_{P_Q}(t) = \left\lceil \frac{avgQJ(t)}{MQL_{UT}} \right\rceil \quad (2)$$

The total number of servers to be provisioned is the sum of $N_P(t)$ and $N_{P_Q}(t)$. Since VM provisioning requires some time, to avoid unnecessary oscillations in the number of VMs, the algorithm adds some delay. In the above discussion, the VM allocation decisions are based on the accumulative play rate and the transcoding rate and are same as mentioned in [11]. However, resource allocation based on accumulative play rate and transcoding rate may have high provisioning cost. It does not take into account the current computation load and the system throughput. The required transcoding rate could be lower due to improper load balancing. Therefore, provisioning more number of servers will result in higher cost. Hence, we also take into account the computation load and the system throughput to provide an even more cost-efficient algorithm for resource allocation.

To determine the number of actual servers required, we first divide the transcoding requests into different classes according to the computation requirements as shown in Figure 2. Subsequently, we check how much throughput can a transcoding server provides for these different classes. For example, there could be following classes of streams according to the video transcoding rate.

i. High Definition (HD) streams with average transcoding rate of 60 frames per second (fps).
ii. Standard Definition (SD) streams with average transcoding rate of 132 fps.
iii. Mobile streams with average transcoding rate of 410 fps

When a new transcoding request arrives in the system, some segments of the stream are sent for transcoding.

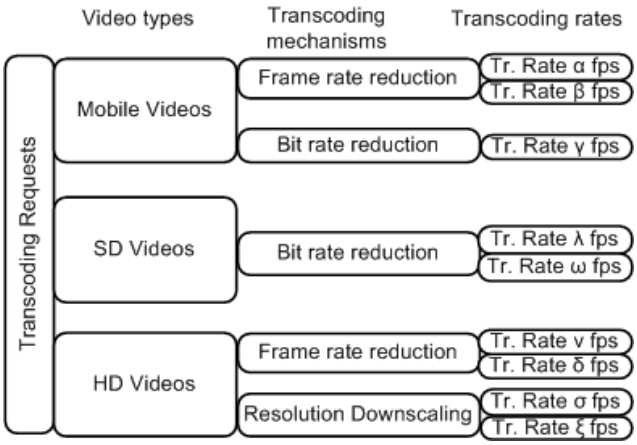


Figure 2: Classification of transcoding requests

Based on the transcoding rate of these segments, the entire video stream is placed in a certain class. The transcoding rate may fluctuate for a video steam. However, if it is within upper and lower bounds of that class, the stream will remain in the same class. In case if the transcoding rate crosses the bounds of that class, it is placed into another class. It means that the runtime class migration of video streams is also possible.

After the classification of video streams, we compute the total number of transcoding requests in each class. For example, there can be 20 requests in the first class, 70 requests in the second class and so on. Since we know the approximate transcoding rate of each class and also the number of streams in each class, we compute the number of servers required.

Let us assume there are n video transcoding requests in the system, denoted by req_1, req_2, ... req_n. These requests can be classified into m different classes C_1, C_2, ..., C_m. The streams in the same class should have similar play rate and similar average transcoding rate. The number of transcoding servers for each class is computed as follows.

$$NSC_k(t) = \left\lceil \frac{NStrmC_k(t) * PRC_k(t)}{avgTRC_k(t)} \right\rceil \quad (3)$$

Finally the total number of actual servers required for all classes of video streams is calculated by the summation of the number of actual servers required for individual classes, as given by equation 4.

$$NAS(t) = \sum_{k=1}^{m} NSC_k(t) \quad (4)$$

We check if the actual number of servers required at this time is less than the number of servers which are already provisioned, then there is need to provision more servers. However, if the number of already provisioned servers is higher than the actual, then there is no need to provision any more servers.

3.2 VM De-allocation Algorithm

The resource de-allocation mechanism is described in Algorithm 2. This algorithm de-allocates virtual machines if the number of provisioned servers $S(t)$ is more than the total number of servers $NAS(t)$ required to perform video transcoding of different classes of video streams and the buffer size $B_S(t)$ exceeds its upper threshold B_U for a predetermined amount of time.

While performing the resource de-allocation, the algorithm computes the remaining renting time of each transcoding server $RT(s,t)$. If the algorithm determines that there are some transocoding servers $S_c(t)$ whose remaining times are in the range of the pre-determined upper threshold of remaining time RT_U and the lower threshold of remaining time RT_L, then it calculates the number of servers to terminate $N_T(t)$ as follows.

$$N_T(t) = \lceil len(S(t)) - NAS(t) \rceil - N_B \quad (5)$$

The $N_T(t)$ servers having lowest remaining renting times are selected for termination. Once a server is selected for termination, it stops accepting new transcoding tasks. As the transcoding time of the jobs is relatively less, the selected servers complete all jobs in queue before termination.

Algorithm 2 VMs de-allocation algorithm

1: **While true do**
2: if $len(S(t)) > NAS(t) \wedge B_S(t) > B_U \wedge count_o(t) > C_T$ then
3: **for** $s \in S(t)$ **do**
4: $RT(s,t) := calRT(s,t)$
5: **end for**
6: $S_c(t) := \{\forall s \in S(t) \mid RT(s,t) < RT_U \wedge RT(s,t) > RT_L\}$
7: **if** $|S_c(t)| \geq 1$ **then**
8: $N_T(t) := calcN_T()$
9: $N_T(t) := min(N_T(t), |S_c(t)|)$
10: **if** $N_T(t) \geq 1$ **then**
11: $sort(S_c(t))$

12: $S_t(t) := select(N_T(t))$
13: $S(t) := S(t) \setminus S_t(t)$
14: $delay(jobCD)$
15: $term(S_t(t))$
16: **end if**
17: **end if**
18: **end if**
19: **end while**

4. SIMULATION RESULTS

In real systems involving complex environment, it is always possible to use software simulations to test and evaluate new algorithms [7]. In this regard, we have used Python programming language to perform simulations. Our discrete-event simulation is based on the SimPy simulation framework [13], which is a library for writing discrete-event simulations in Python.

4.1 Experimental Design and Setup

We use a computation cost model similar to Amazon EC2 in which the renting period of a VM is based on an hourly charge model. In our experiments, we use only small instances and its cost is assumed to be $0.06 per hour. The storage cost is based on a nonlinear model similar to the Amazon S3 cost model in which for the first Tera Byte (TB), the renting cost is $0.095 per Giga Byte (GB) per month. For the next 49 TB, the cost is $0.080 per GB per month. In the similar way, for next 450 TB, the cost is $0.070 per GB per month. For next 500 TB, it is $0.065 per GB, for next 4000 TB, it is $0.060 per GB and for over 5000 TB, the storage cost is $0.055 per GB per month. We use three different types of videos namely HD, SD and mobile video streams. Mostly SD videos are used for streaming over the internet. Therefore, we considered 50% SD videos, 20% HD videos and 30% mobile videos. The transcoding time of a GOP for different types of video varies. Therefore, the average size of a segment for HD video is 68 to 82 frames, while for mobile video and SD videos, it is from 230 frames to 270 frames. We consider two different load patterns in two separate experiments. Load pattern 1 in experiment 1 consists of 30 possible video formats for the transcoded videos. Whereas, load pattern 2 in experiment 2 has 45 possible video formats for the transcoded videos. The upper threshold RT_U of the remaining time is taked as 60 seconds, while the lower threshold RT_L of the remaining time is taked as 12 seconds. The total number of frames in a video stream is taken in the range of 15000 to 18000. The desired play rate for a video stream is often kept fixed, i.e., 30 fps for SD videos and mobile video streams, while 24 fps for HD videos. The transcoding rate for HD videos is assumed to be between 36 fps to 48 fps. For SD and mobile videos it is assumed to be 60 fps to 120 fps.

4.2 Results

In Figures 3, 4, 5, and 6 the plots of user requests show the number of video transcoding requests in the system. The cost plot shows the transcoding as well as the storage cost in US dollars. The videos plots indicate the number of source videos and transcoded videos. The load patterns used in both experiments are synthetic, involving two peaks and minor fluctuations in the load. Each transcoded video is also stored in the video repository. Therefore, for subsequent requests for the same transcoded video, it is not needed to perform the transcoding operation. We performed two experiments as follows.

1) Experiment 1 - 30 possible transcoded formats:

Figure 3 shows results from experiment 1. This experiments uses the algorithms described in [11]. In this experiment, a maximum of 212 transcoding servers are used. The average number of servers is 96 and the minimum number of servers is 2. The two-days renting cost of transcoding servers is 220 dollars and the storage cost is 180 dollars.

Figure 4 shows the results from experiment 1 using the improved algorithms proposed in this paper. In this experiment, a maximum of 114 transcoding servers are used. The average number of servers is 42 and the minimum number of servers is 2. The two-days renting cost of transcoding servers is 103 dollars and the storage cost is 180 dollars. Therefore, the results indicate that the storage cost for old and new algorithms is same as expected. Whereas, the transcoded cost with the improved algorithms is less than half of the old algorithms for two days of simulation results. Hence, it is evident that the existing algorithms take an aggressive decision while provisioning the servers and result in higher trasncoding cost.

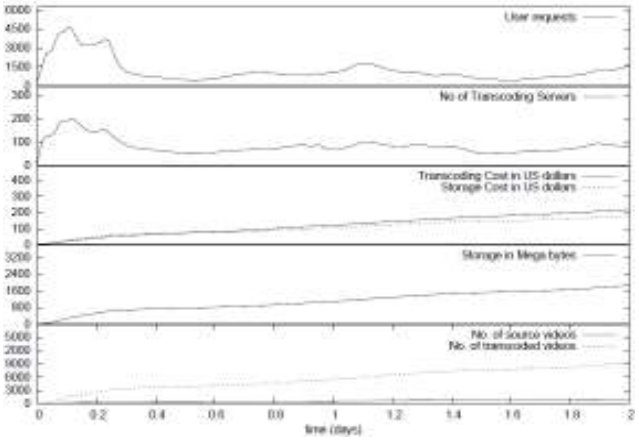


Figure 3: Experiment 1 results of old approach

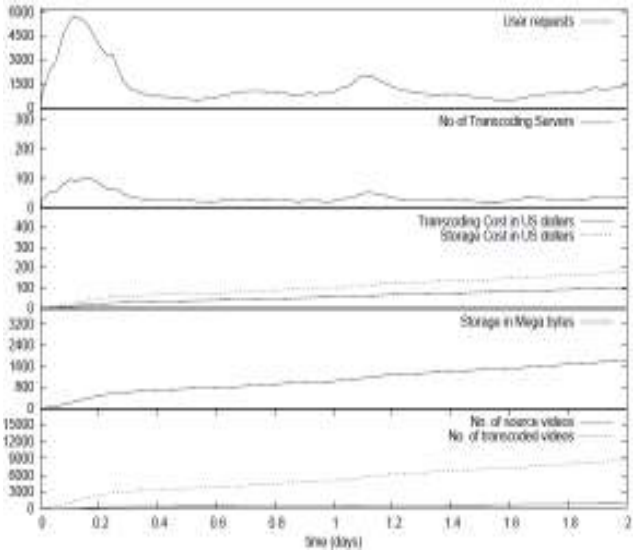


Figure 4: Experiment 1 results of new approach

2) Experiment 2 - 45 possible transcoded formats:

Figure 5 shows results from experiment 2. In this experiment, the possible number of transcoded videos for each source video is 45 which requires more transcoding operations as compared with the results in experiment 1. This experiment also uses the same algorithms as proposed in [11]. In this experiment, a maximum of 304 transcoding servers are used. The average number of servers is 138 and the minimum number of servers is 2. The two-days renting cost of transcoding servers is 301 dollars and the storage cost is 270 dollars. Figure 6 shows the results from experiment 2 using the improved algorithms proposed in this paper. In this experiment, a maximum of 182 transcoding servers are used. The average number of servers is 65 and the minimum number of servers is 2. The two-days renting cost of transcoding servers is 149 dollars and the storage cost is 270 dollars.

From the experiments, it is evident that the new algorithms have approximately half cost as compared with the previous algorithms.

5. CONCLUSION

In this paper, we presented virtual machine allocation algorithms for video transcoding in an IaaS cloud. The proposed algorithms are also compared with the existing, prediction based video transcoding algorithms in a cloud computing environment.

The proposed algorithms use accumulative play rate, accumulative transcoding rate, computation load and system throughput to take resource allocation and deallocation decisions. The proposed algorithms also use a two-step prediction method to predict the future load in order to allow proactive resource allocation. We used GOP level video segmentation to distribute transcoding load among different transcoding servers. The results in both experiments (exp.1 and exp.2) indicate that the proposed algorithms are more cost-efficient as compared to the existing algorithms. In both the experiments, the transcoding cost with the proposed algorithms is almost half as compared to the transcoding cost with traditional algorithms [11].

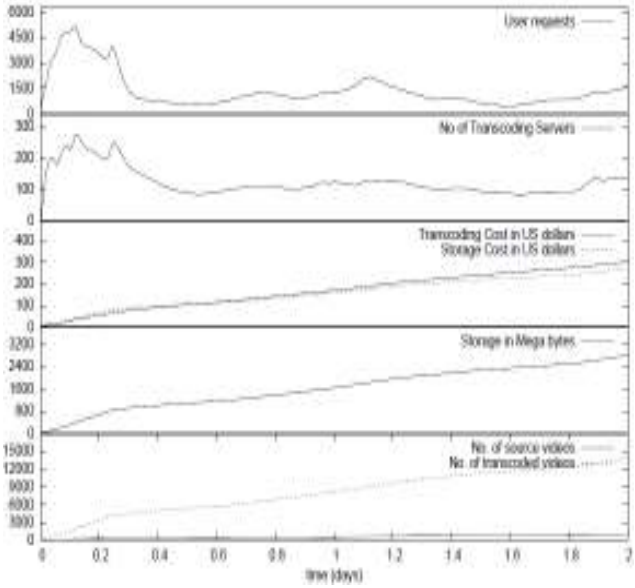


Figure 5: Experiment 2 old results

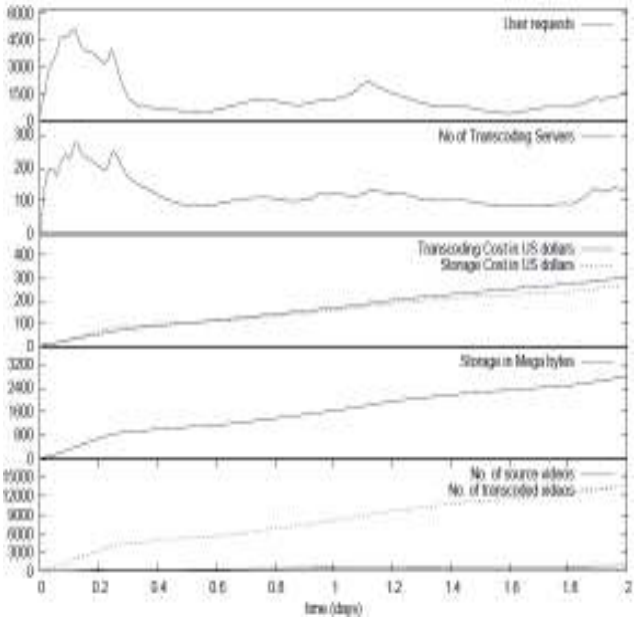


Figure 6: Experiment 2 new results

REFERENCES

[1] Mauro Andreolini, Sara Casolari, and Michele Colajanni. Models and framework for supporting runtime decisions in web-based systems. ACM Trans. Web, 2(3):17:1–17:43, July 2008.

[2] Mauro Andreolini and Sara Casolari. Load prediction models in web-based systems. In Proceedings of the 1st international conference on Performance evaluation methodolgies and tools, valuetools '06, New York, NY, USA, 2006. ACM.

[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A view of cloud computing. Commun. ACM, 53(4):50–58, April 2010.

[4] Adnan Ashraf, Benjamin Byholm, Joonas Lehtinen, and Ivan Porres. Feedback control algorithms to deploy and scale multiple web applications per virtual machine. In Software Engineering and Advanced Applications (SEAA), 2012 38th EUROMICRO Conference on, pages 431 –438, sept. 2012.

[5] Adnan Ashraf, Fareed Jokhio, Tewodros Deneke, Sebastien Lafond, Ivan Porres, and Johan Lilius. Stream-based admission control and scheduling for video transcoding in cloud computing. in Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on, pages 482–489.

[6] N. Bjork and C. Christopoulos. Transcoder architectures for video coding. Consumer Electronics, IEEE Transactions on, 44(1):88 –98, feb 1998.

[7] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, C. A. F. De Rose, and Rajkumar Buyya. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41(1):23–50, 2011.

[8] Waheed Iqbal, Matthew N. Dailey, David Carrera, and Paul Janecek. Adaptive resource provisioning for read intensive multi-tier applications in the cloud. Future Generation Computer Systems, 27(6):871 – 879, 2011.

[9] Fareed Ahmed Jokhio, Tewodros Deneke, Sébastien Lafond, and Johan Lilius. Analysis of video segmentation for spatial resolution reduction video transcoding. In Intelligent Signal Processing and Communications Systems (ISPACS), 2011 International Symposiuml, page 6 pp., Dec 2011.

[10] Fareed Jokhio, Adnan Ashraf, Sébastien Lafond, and Johan Lilius. A computation and storage trade-off strategy for cost-efficient video transcoding in the cloud. In 39th EUROMICRO Conference on Software Engineering and Advanced Applications, pages 365–372. IEEE Computer Society, 2013.

[11] Fareed Jokhio, Adnan Ashraf, Sebastien Lafond, Ivan Porres, and Johan Lilius. Prediction-based dynamic resource allocation for video transcoding in cloud computing. In Parallel, Distributed and Network-Based Processing (PDP), 21st Euromicro International Conference on, pages 254–261, 2013.

[12] F. Jokhio, T. Deneke, S. Lafond, and J. Lilius. Bit rate reduction video transcoding with distributed computing. In Parallel, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on, pages 206 –212, feb. 2012.

[13] Norman Matloff. A Discrete-Event Simulation Course Based on the SimPy Language. University of California at Davis, 2006.

[14] J. Rhoton and R. Haukioja. Cloud Computing Architected: Solution Design Handbook. Recursive Press, 2011.

[15] J. Rhoton. Cloud Computing Explained. Recursive Press, 2010.

[16] A. Vetro, C. Christopoulos, and Huifang Sun. Video transcoding architectures and techniques: an overview. Signal Processing Magazine, IEEE, 20(2):18 – 29, mar 2003.

[17] J. Watkinson. The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4. Broadcasting and communications. Elsevier/Focal Press, 2004.

[18] T. Wiegand, G. J. Sullivan, and A. Luthra. Draft ITU-T recommendation and final draft international standard of joint video specification. In Technical Report, 2003.

[19] Andreas Wolke and Gerhard Meixner. TwoSpot: A cloud platform for scaling out web applications dynamically. In Elisabetta Di Nitto and Ramin Yahyapour, editors, Towards a Service-Based Internet, volume 6481 of Lecture Notes in Computer Science, pages 13–24. Springer Berlin / Heidelberg, 2010.

COMPARATIVE STUDY OF STATIC TORQUE CHARACTERISTIC OF SWITCHED RELUCTANCE MOTOR USING SPLINE AND LINEAR DATA INTERPOLATION TECHNIQUES

Rameez Akbar Talani*, Ghulam Mustafa Bhutto*, Fareed Hussain Mangi**, Muhammad Usman Keerio*

ABSTRACT

This Paper presents the comparative study of static torque characteristic of switched reluctance motor under spline and linear data interpolation technique. Simulation model for static torque characteristics of switched reluctance motor is presented using spline and linear data interpolation technique to achieve the best cure fit technique. The findings of this study are helpful to achieve the best simulation model, which helps to predict the performance of motor. Due to highly nonlinear curves and big data tables, it is very difficult to predict the performance of switched reluctance motor. Therefore it is always required to choose the suitable curve fit technique. The results obtained in this paper will help in accurate modeling of switched reluctance motor.

Keywords: *Static Torque Characteristics, MATLAB, Switched Reluctance Machine.*

1. INTRODUCTION

Switched reluctance motor has salient type construction i.e. salient poles on rotor and stator. Stator carries the main or field winding whereas rotor does not carries any winding. The semiconductor switches are connected in series with windings of the machine to form a phase. Switched reluctance machine is a multi-phase type machine. SR machine have different number of stator and rotor poles [1].

Due to salient type construction, switched reluctance machine has non-linear nature and magnetic saturation. In switched reluctance motor position dependent switching are key factor, which are achieved by proper electronic control [1]

Appropriate switching and conduction angles are necessary for proper operation of the switched reluctance motor. The switch gate signals with respect to the rotor position are applied at the gate-emitter terminal of the switching devices which are used in the converter of the motor.

The duration of gate signals for which the switching device remains in ON state is known as conduction angle [4].

In switched reluctance machine, rotation is due to magnetic pull of stator poles to rotor poles. Rotor poles rotate from maximum reluctance region to minimum reluctance region where as low region to high region for inductance [2].

Machine is said to be in aligned position when stator and rotor poles are complete overlapped and is said to be unaligned position when rotor and stator poles are not overlapping to each other [3].

Due to highly nonlinear nature of switched reluctance motor, it is very necessary to achieve the best curve fit technique. The literature presented in this paper is mainly concerned with impact of different curve fitting techniques on modeling of switched reluctance motor.

In paper [1] simulation model for sr motor is presented using a suitable curve fit. Polynomial curve fit technique is used for magnetic modelling of machine.

In paper [5] nonlinear behavior of SR motor is discussed. 2D bi cubic spline method is implemented to curve fit. Calculated and experimental results are than compared to measure the degree of accuracy.

Cubic spline curve fitting technique is implemented in paper [6,7]. It is shown that smooth curves are obtained using spline. Paper [8] shows the static torque characteristics of switched reluctance motor using appropriate curve fit. A bi-cubic spline interpolation method is implemented to achieve smooth curves in paper [10]. These curves show that phase flux linkage depends on the phase current at different rotor positions and are obtained experimentally.

Literature presented in [1], [5-10] highlight application of any one curve fitting technique when applied on flux linkage characteristics or static torque characteristics and

* Electrical Engineering Department,

** Energy and Environment Engineering Department,

Quaid-e-Awam University of Engineering, Science and Technology Nawabshah Pakistan.

have not applied the comparative study to achieve the best curve fit. This paper mainly focus on spline and linear data interpolation techniques when applied on static torque curve to achieve the best curve fit technique.

2. STATIC TORQUE CHARACTERISTICS

Static torque curves are function of rotor position [8]. Fig.[2.1] shows static torque curve from un aligned to aligned rotor position. Stator and rotor pair of poles are completely unaligned at -30 angle, as angle decreases poles starts to be aligned and 0 angle is the position when stator and rotor pair of poles are completely aligned. From Fig.[2.1] it can be seen that, near to -15 angle is the position when maximum torque is achieved and no or zero torque is produced when stator and rotor pair of poles are completely aligned i.e. at 0 angle position

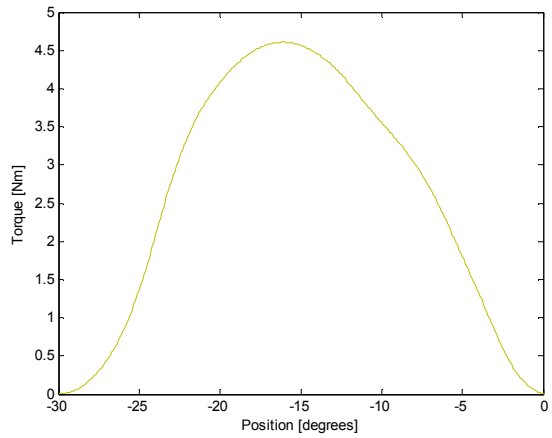


Figure: 2.1 Torque versus rotor position

The data of static torque is calculated once the data of co-energy is known using equation (1)

$$T(\theta, i) = \frac{\delta \omega(\theta, i)}{\delta \theta} \quad (1)$$

Where angle θ is rotor position and i is current carrying coil. Co-energy is represented by ω which is the function of rotor position and coil current.

3. SIMULATION MODEL AND COMPARISION

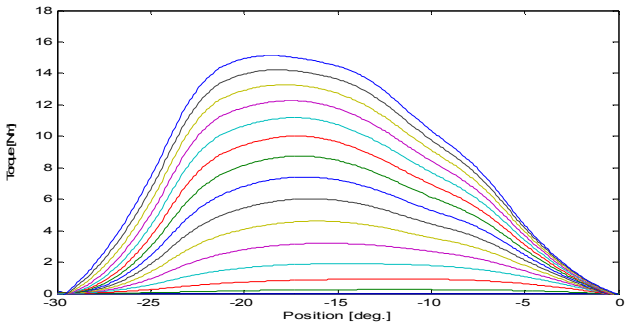


Figure 3.1 Torque versus rotor position [under spline]

Fig.[3.1] shows static torque characteristics which are function of rotor position. Curves are obtained under spline data interpolation technique. From above figure it can be seen that smooth curves are achieved under spline curve fit.

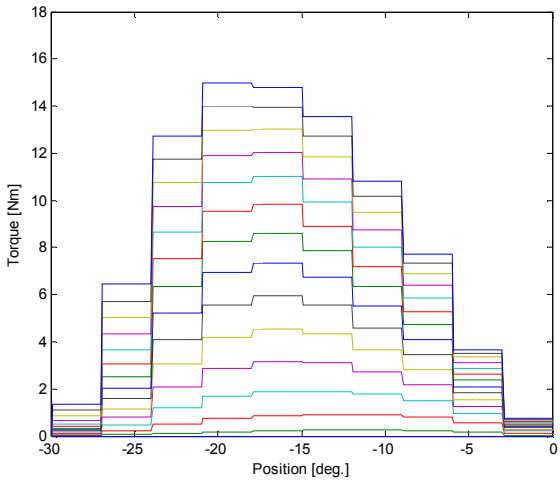


Figure 3.2 Torque versus rotor position [under linear]

Fig.[3.2] shows static torque curves under linear data interpolation technique. Fig.[3.2] has zero bulges and lot of errors. From fig.[3.1] and fig.[3.2] it is shown that static torque curves obtained under spline curve fit are smooth and has no errors whereas as the curves produce under linear are not smooth and has errors.

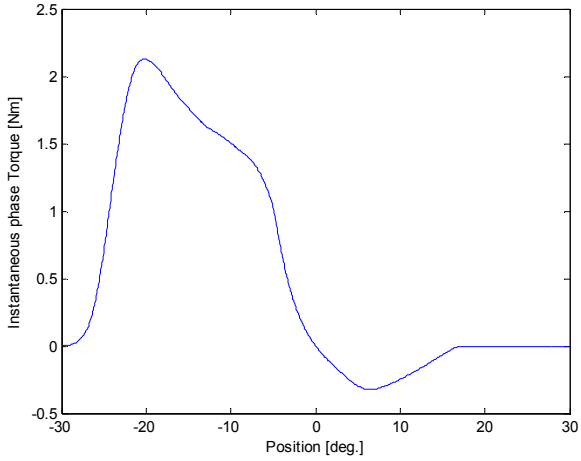


Figure 3.3 Instantaneous Phase Torque versus rotor position [under spline]

Fig 3.3 shows the instantaneous phase torque versus rotor position curve using cubic spline curve fit. From above fig it seems that as the rotor pair of poles changes its position from unaligned to aligned (-30 deg to 0 deg) the instantaneous value of torque increases sharply from -30

deg to -20 deg and starts decreasing from -20 deg to 0 deg. The curve is seems to be ok

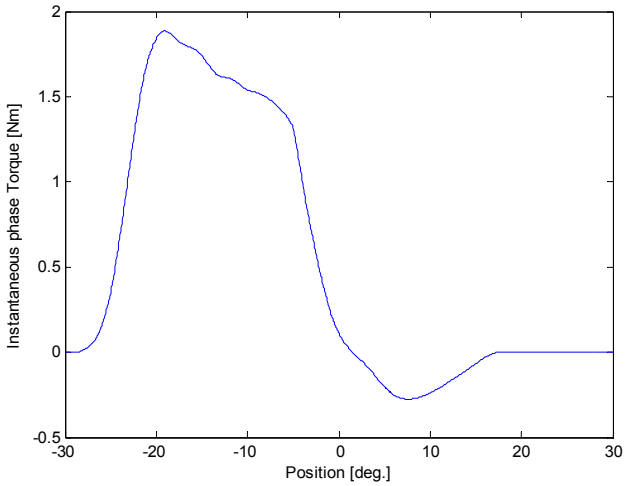


Figure 3.4 Instantaneous Phase Torque versus rotor position [under linear]

When comparing the Fig 3.4 (under spline) with the Fig 3.3 (under linear) it seems that the curve under linear interpolation has some distortion between -20° to -5° deg. From above fig it seems that as the rotor pair of poles changes its position from unaligned to aligned (-30 deg to 0 deg) the instantaneous value of torque increases sharply from -30 deg to -20 deg and starts decreasing from -20 deg to 0 deg. At -5 degree the switch turns off but torque does not decrease instantly because of current, which does not falls rapidly due to R L circuit.

4. CONCLUSION

Static torque characteristic that is the highly nonlinear function of torque and rotor position is achieved using curve fit techniques. Cubic spline data interpolation and linear data interpolation techniques are employed and compared to achieve the best curve fit. From above results it is concluded that smooth and regular curve is achieved using spline curve fit technique while the linear has errors. Results are helpful in performance prediction of machine.

REFERENCES

[1] S.Khotpanya, S.Kittiratsathca, Ishi Kazuhisa, "A Magnetic Model of a Three-Phase Switched Reluctance Machine using Cubic Spline Interpolation Technique", Presented at Power Electronics and Drives Systems PEDS, International Conference on Vol..2, pp.1167 – 1170, 2005.

[2] Liuchen Chang, "Modelling of Switched Reluctance Motors", Electrical and Computer Engineering, Engineering Innovation: Voyage of Discovery. IEEECanadian Conference on, Vol.2, pp.866- 869, 1997.

[3] Wei Min Chan and W. F. Weldon, "Development of a Simple Nonlinear Switched Reluctance Motor Model using Measured Flux Linkage Data and Curve Fit", Presented at IEEE Industry Applications Society Annual Meeting New Orleans, Louisiana, October 5-9,1997.

[4] T. Jahan. M.B.B. Sharifian, M.R. Feyzi, "Static Characteristics of Switched Reluctance Motor 6/4 by Finite Element Analysis", Australian Journal of Basic &Applied Sciences, pp. 1403-1411, September 2011.

[5] Chang yan tai, Cheng k. w. eric, "A Simulation Model for a 4 Phase Switched Reluctance Motor For PSIM", Presented on Power Electronics Systems and Applications (PESA), 4th International Conference on, 2011.

[6] D.W.J. Pulle,"New data base for switched reluctance drive Simulation", IEE PROCEEDINGS-B, Vol. 138, No. 6, pp.331-337, November.

[7] ShileiXu , Desheng Li, Gaofei Deng, Guotian Chen, "A Single-Phase Switched Reluctance Motor Drive System", Power Engineering and Automation Conference IEEE, pp. 264- 267 , 8-9 September 2011.

[8] V. Ramanarayanan, L.Venkatesha, Debi prasad Panda, "Flux Linkage Characteristics of Switched Reluctance Motor", Power Electronics, IEEE Proceedings of the 1996 International Conference, pp. 281 – 285, 8-11 Jan 1996.

[9] P.J.Lawrenson, J.M.Stephenson, P.T.BIenkinsop, J.Corda, N.N.Fulton, "Variable-speed switched reluctance motors", IEE PROC, Vol. 127, Pt. B, No. 4, pp.253-265, JULY 1980

[10] Xiang-Dang Xue, K. W. E. Cheng, S. L. Ho, "Simulation of Switched Reluctance Motor Drives Using Two-Dimensional Bicubic Spline", IEEE Transactions on Energy Conversion, pp. 471-477, December 2002.

SURFACE ADSORPTION STUDY OF SAPONIFIED ORANGE WASTE GEL FOR ARSENIC (III) REMOVAL

Khadija Qureshi*, Yasir Hameed Mangi**, Fareed Hussain Mangi***,

Kashif Hussain Mangi****, Saleem Raza Samo***

ABSTRACT

Arsenic is mostly present in Sindh and Punjab provinces of Pakistan. In some districts of Punjab and Sindh arsenic concentration has reached up to the level of 906 ppb and 200 ppb respectively. Since 1993, the World Health Organization has recommended a maximum contaminant level for arsenic in drinking water of 10 ppb (except for a few countries) and the water treatment system must comply with this standard. Orange juice waste was used to prepare adsorbent gel by loading iron oxide. The iron oxide loaded SOW gel was used to remove As (III) from water. Batch experiments conditions were conducted to optimize the adsorbent dose, shaking time, adsorbate concentration for maximum 96% removal of arsenic (III) from aqueous solution.

1. INTRODUCTION

The rapid growth of population in Pakistan the available water resources are about to exhaust and there is strong need to treat the underground water making it safe for use. Arsenic is mostly present in Sindh and Punjab provinces of Pakistan. Pakistan council for research and water resources (PCRWR) conducted survey under UNICEF arsenic monitoring program in 2005, it was found that ground water collected from 60 locations in four districts of Sindh contained arsenic from 100 to 500 ppb [1]. Arsenic problem is also identified as a result of field testing in ground water of various districts of Punjab like Attock and Rawalpindi (year 2000), and also in Muzzafargarh concentrations of As reached up to 906 ppb Arsenic (As). It was reported that arsenic contamination exceeded 200 ppb in some districts of Sindh like Khairpur, Dadu, Nawabshah, Nausheroferoz, and Thatta [1], and 170 ppb in Tando Allahyar. It was identified as one of the major issue in some areas of both provinces of Pakistan i.e. Southern Punjab and Central Sindh. It is one of the most toxic contaminants found in the environment and can exist as various complex forms or species in the aquatic environment. Arsenite (AsO_3^{3-}) and arsenate (AsO_4^{3-}), referred to as As (III) and As (V), respectively, are common in natural waters [7]. The presence of arsenic in ground water has been reported in many countries like Bangladesh China Iran Nepal Pakistan hence it is a global issue. In Pakistan the underground and tube wells are contaminated with arsenic reported to be above the recommended USEPA arsenic level of 10ppb. In order to overcome the drawbacks of these traditional treatment processes, attempts have been made regarding the removal of As (V) and As (III) by using Fe (III)-loaded chelating ion exchange resins having either an acidic or basic moiety as the functional group. However, treatment with such resins is rather expensive for the recovery of arsenic. At this juncture adsorption of arsenic using natural products and biomass has emerged as an option for developing economic and eco-friendly wastewater treatment processes [5]. Pakistan is the largest producer of 'Citrus Reticula' variety (Kinow) [2], this unique variety of citrus is indigenous to this part of the world. In the present research the Natural adsorbent was prepared from orange waste, which is totally bio nature and safe for human being, as well as more effective and economic for our country because Pakistan produces about 133600 tons of mandarins orange predominantly Kinnow. Which is about 95% of world kinnow production [2].

2. MATERIAL AND METHODS

The first phase of work was preparation of saponified orange waste gel (SOW gel) from orange juice waste along with loading of iron oxide. The second phase of the work was optimization of adsorption parameters such as adsorbent dose, shaking time, and adsorbate concentration for the use of SOW gel for the removal of Arsenic from aqueous solution.

* Department of Chemical Engineering, Mehran UET Jamshoro.

** OMV Pakistan Limited.

*** Department of Energy & Environment Engineering, QUEST Nawabshah.

**** Department of Chemical Engineering, University of Karachi.

Fig-1: Schematic of the Experimental Benchmark

2.1 Raw Material Collection

Maximum percentage of orange production (70%) is utilized to produce derivative products and approximately 50–60% of the processed fruit is transformed into citrus peel waste i.e. peel, seeds and membrane residues (Wilkins et al., 2007a). Approximately 20.17% represents the solid material existence in the orange fruit which has a potential to be used as adsorbent for arsenic removal converting approximately 4 million tons of orange waste into useful product and eliminating an environmental waste. The orange juice waste was collected from local juice shop, and iron oxide was purchased of analytical grade from local market.

2.2 Preparation of Saponified Orange Waste Gel

The orange juice waste was treated with calcium hydro oxide to carry out the saponification reaction to form saponified orange waste gel as shown in fig: 2a.

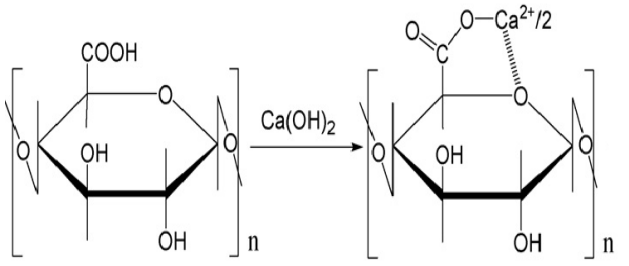


Fig-2a: Saponification of orange juice waste

In fig: 2a H^+ in the –COOH functional group has been replaced by calcium forming saponified orange waste and yielding H_2O as by product. In order to carry out the reaction orange waste was processed through various unit operations including crushing, washing, drying and mixing. Series of operations were carried out including Mixing, washing, drying and grinding etc. to produce SOW gel for arsenic removal.

The SOW gel achieved was repeatedly washed with deionized water by means of decantation process until neutral pH; acquired Wet gel then shifted into a china dish and left for 48 hours in a forced convection oven (FC-42D) at 70 C to produce dry SOW gel. Dried Gel further grinded in a pistil morter up to the favorable size for the loading process.

2.3 Loading of Iron oxide (Fe_2O_3) on SOW Gel

SOW gel being a cation exchanger cannot adsorb arsenic species and it is necessary to load metal ion [19]. The extent of adsorption largely depends on the kind of metal ion loaded onto the gel, moreover in aqueous solution, iron gives $[Fe(OH)(H_2O)5]^{2+}$ ion [26], which facilitates ligand exchange and can alter the properties of SOW gel to be suitable for the arsenic removal [21]. The loading of metal ion on the saponified orange waste gel results in the elimination of calcium and it is replaced by metal ion suitable for the ligand exchange.

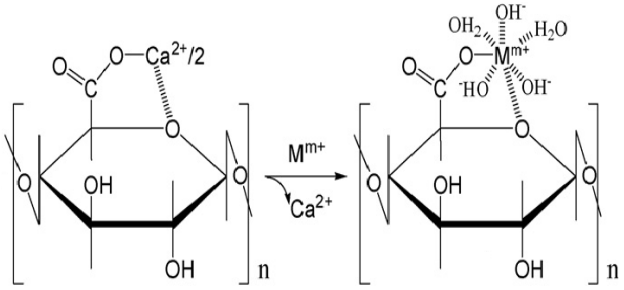


Fig-2 b: Loading of iron oxide (Fe_2O_3) on SOW gel

In order to load iron oxide onto the saponified orange waste SOW gel 3 grams of the SOW gel were equilibrated with 500 ml of 0.1M iron oxide solution at pH 2 for 24 hours and then filtered from iron oxide solution by a filter paper. Subsequently gel was achieved after filtration was characterized as iron oxide (Fe_2O_3) loaded SOW gel**.** The (Fe_2O_3) loaded SOW gel obtained was than repeatedly washed with deionized water until neutral pH. After neutralization of iron oxide loaded saponified orange waste gel, it was dried in (FC-42D) oven until constant weight, the temperature of oven was raised to 110 C. The surface area of the adsorbent for the removal of arsenic from aqueous solutions was increased by further grinding of dried (Fe_2O_3) loaded SOW gel. The grinded (Fe_2O_3) loaded SOW gel was sieved to obtain a uniform particle size fraction of between 75 and 150 µm for the adsorption tests.

2.4 Adsorption of Arsenic on Fe2O3 Loaded SOW Gel

Adsorption of arsenic onto Fe_2O_3 loaded SOW gel is governed by ligand exchange as shown in fig-3.

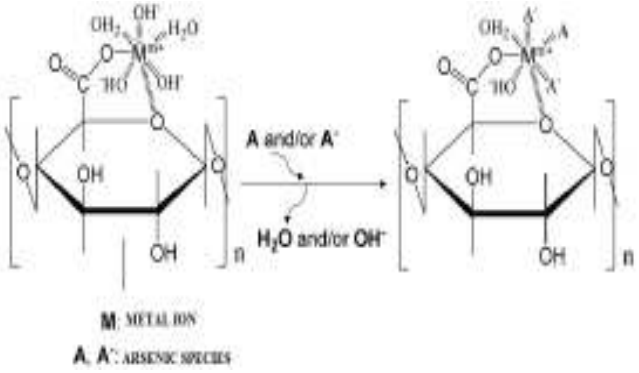


Fig-3: Adsorption of arsenic on the (Fe_2O_3) loaded SOW gel

Batch-wise adsorption tests for arsenic were carried out to examine the adsorption behavior of arsenic on the (Fe_2O_3) loaded SOW gel.

3. RESULTS AND DISCUSSION

The adsorption tests were carried out in the concentration range of 1ppm to 10 ppm. First of all adsorption of arsenic as a function of adsorbent concentration (adsorbent dose) was examined in a series of experiments where the initial concentration of arsenic was maintained constant at 1 ppm (mg/l) and thus the optimum adsorbent concentration was determined. These experiments were carried out at pH 10 on the basis of previous studies. All batch adsorption experiments were carried out in a 50 ml conical flask by taking 15 ml of arsenic solution and varying concentration of adsorbent. Than these samples were put in to a thermo shaker at 180 rpm for 100 minutes and then filtered to investigate the removal efficiency. Further operating parameters such as adsorbent dose(0.25g, 0.5g, 0.75g, 1g, 1.25g) shaking time (25 min, 50 min, 100 min, 125min) and metal concentration (1ppm, 5ppm, 10ppm, 20ppm) were optimized through batch type adsorption tests by keeping following parameters constant Temperature, pH, and shaking speed.

3.1 Optimization of Operating Parameters

In order to optimize operating parameters batch type adsorption tests were carried out by keeping following parameters constant Temperature, pH, shaking speed and adsorption parameters such as adsorbent dose(0.25g, 0.5g, 0.75g, 1g, 1.25g) shaking time (25 min, 50 min, 100 min, 125 min) and metal concentration (1ppm, 5ppm, 10ppm, 20ppm) were investigated.

3.2 Effect of adsorbent dose on Arsenic III removal

Batch type experiments were carried out in a 50 ml conical flask by keeping following parameters constant Temperature 30 ^{0}C, pH10, shaking speed 180 rpm, shaking time 100 minutes and adsorbate concentration 1ppm in order to examine the sorption efficiency of the Fe_2O_3 Loaded SOW gel for the removal of arsenic as a function of adsorbent dose . fig: 5.1 shows the sorption efficiency of Fe_2O_3 loaded SOW gel, maximum sorption efficiency is achieved at 1 gram adsorbent dose which is 86% removal efficiency and at 0.75 gram adsorbent dose sorption efficiency is 85% which is almost same for the 1 gram dose hence the optimum dose for the batch experiments was decided as 0.75 gram adsorbent which is about 50 g/L.

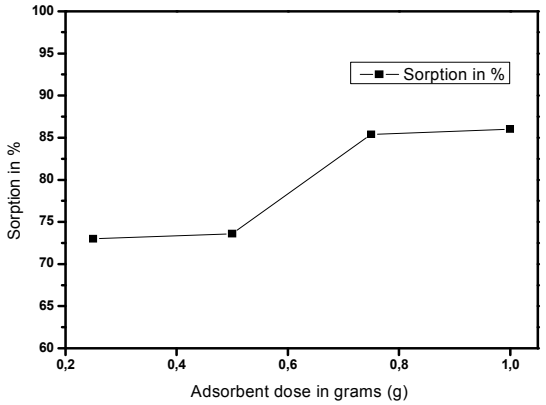


Fig-4: Effect of adsorbent dose on Arsenic III removal

In previous research adsorbent dose of 0.25 gm have been reported while using red mud as an adsorbent [27]. However more than 70% removal can be achieved at 17 g/L adsorbent dose.

3.3 Effect of shaking time on Arsenic III removal

Fig: 5 shows the optimum time required for the removal of arsenic III by Fe_2O_3 Loaded SOW gel by keeping 0.75 gram of adsorbent dose and all other parameters like temperature pH and rpm constant.

Different shaking times ranging from 0 min to 150 min were investigated as a function of arsenic removal

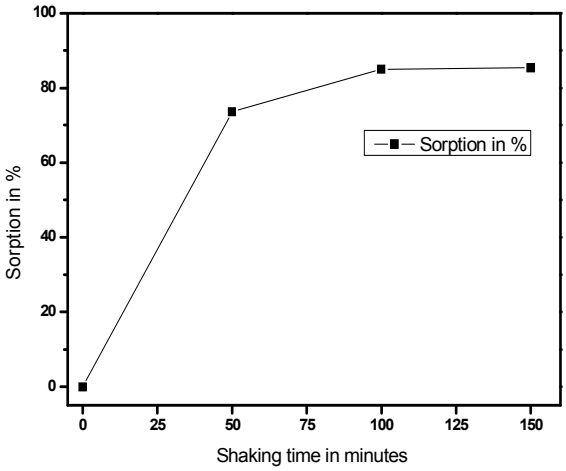


Fig-5: Effect of shaking time on Arsenic III removal

efficiency from aqueous medium. The maximum removal efficiency is achieved at 150 minutes which is 85.4% arsenic removal although a significant increase in sorption efficiency is observed from 79.6% to 85% in the shaking time ranging from 50 minutes to 100 minutes.

After 100 minutes the increase in sorption efficiency is insignificant hence the shaking time is optimized at 100 minutes for the batch experiments to remove arsenic from aqueous solution which is less when compared with previous research 92 % efficiency was achieved at 240 minutes [28].

3.4 Effect of adsorbate concentration on Arsenic removal

Different adsorbate concentrations were treated with Fe_2O_3 Loaded SOW gel in order to examine the efficiency of adsorbent in high concentration solutions. All the parameters including temperature, pH, shaking speed were kept constant as in earlier experiments and the remaining optimized parameters were maintained however adsorbate concentration used was in the range of 1 ppm to 20 ppm, as reported in previous studies [29, 30]. Above figure shows the efficiency of the Fe_2O_3 Loaded

SOW gel increases with increase in adsorbate concentration it reaches up to the maximum sorption efficiency of 96% at 20 ppm and more than 85% arsenic removal is possible at adsorbate concentration of 1 ppm,

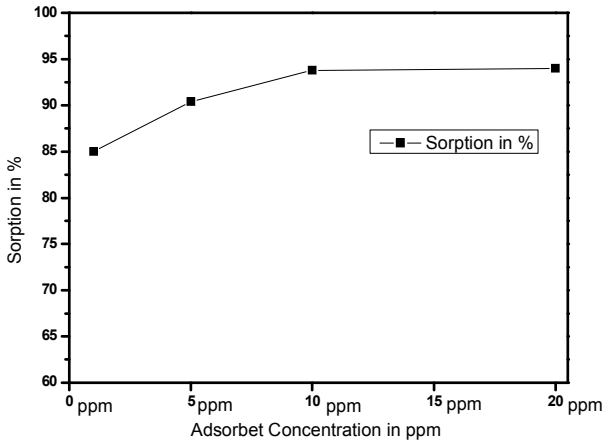


Fig-6: Effect of adsorbate concentration on Arsenic removal

Moreover it also provides knowledge that the same gel could be used repeatedly for the adsorption of arsenic onto it in a low adsorbate concentration.

4. CONCLUSION

The Iron loaded saponified orange waste gel has an intensive potential to replace the presently used expensive methods for removal of arsenic and other heavy metals from ground water in Pakistan. The optimum conditions at which maximum removal efficiency was obtained was 96% at adsorbent dose 0.75 g/L, shaking time 100 minutes, adsorbate concentration 20 ppm. It was found that in alkaline conditions the maximum arsenic removal was achieved. Furthermore Fe_2O_3 loaded saponified orange waste gel can be implied on continuous scale for commercial water filters to remove arsenic.

REFERENCES

[1] PHED Sindh and UNICEF report 2002

[2] http://www.dpi.nsw.gov.au/research/updates/issues/june-2007/citrus-in-pakistan

[3] M.A WAHEED , A Modified Routine Analysis of Arsenic Content in Drinking-water in Bangladesh by Hydride Generation-Atomic Absorption Spectrophotometry J HEALTH POPUL NUTR 2006 Mar, 24(1):36-41 2006 ICDDR, Centre for Health and Population Research.

[4] Biplob Kumar Biswas, Adsorptive removal of As(V) and As(III) from water by a Zr (IV)-loaded orange waste gel, Journal of Hazardous Materials 154 (2008).

[5] C.T. Kamala, Removal of arsenic(III) from aqueous solutions using fresh and immobilized plant biomass, Water Research 39 (2005).

[6] Chia-Pin Chioa, Low-cost farmed shrimp shells could remove arsenic from solutions kinetically, Journal of Hazardous Materials 171 (2009) 859–864.

[7] Yaping Zhao, Synthesis of the cotton cellulose based Fe(III)-loaded adsorbent for arsenic(V) removal from drinking water, Desalination 249 (2009).

[8] Laurent Dambies, Treatment of arsenic-containing solutions using chitosan derivatives: uptake mechanism and sorption performances, Water Research 36 (2002).

[9] H. Soner Altundog, Arsenic removal from aqueous solutions by adsorption on red mud, Waste Management 20 (2000) 761-767.

[10] Diana Q.L. Oliveira, Removal of As(V) and Cr(VI) from aqueous solutions using solid waste from leather industry, Journal of Hazardous Materials 151 (2008) 280–284.

[11] Sanjoy Kumar Maji, Arsenic removal from real-life groundwater by adsorption on laterite soil, Journal of Hazardous Materials 151 (2008) 811–820.

[12] Marı´a P. Elizalde-Gonza´lez, Chemically modified maize cobs waste with enhanced adsorption properties upon methyl orange and arsenic, Bioresource Technology 99 (2008).

[13] P.A.L. Pereira, Adsorptive removal of arsenic from river waters using pisolite , Minerals Engineering 20 (2007) 52–59.

[14] Biplob K. Biswas, Removal and recovery of phosphorus from water by means of adsorption onto orange waste gel loaded with zirconium, Bioresource Technology 99 (2008).

[15] FerideUlua, Removal and recovery of phosphorus from water by means of adsorption onto orange waste gel loaded with zirconium, Separation and Purification Technology (2010).

[16] Kedar Nath Ghimire, Adsorptive separation of arsenate and arsenite anions from aqueous medium byusing orange waste, Water Research 37 (2003).

[17] Ravi Kumar, Removal of As(V) from water by pectin based active hydrogels following geochemical approach, Bioresource Technology 100 (2009).

[18] Kaushik Gupta, Arsenic removal using hydrous nanostructure iron(III)–titanium(IV) binary mixed oxide from aqueous solution, Journal of Hazardous Materials 161 (2009) 884–892.

[19] Chia-Pin Chioa, Low-cost farmed shrimp shells could remove arsenic from solutions kinetically, Journal of Hazardous Materials 171 (2009) 859–864.

[20] Zhijian Li**,** As(V) and As(III) removal from water by a Ce–Ti oxide adsorbent: Behavior and mechanism, Chemical Engineering Journal 161 (2010) 106–113.

[21] Ahmet Sarı Equilibrium, thermodynamic and kinetic investigations on biosorption of arsenic from aqueous solution byalgae (Maugeotia genuflexa) biomass, Chemical Engineering Journal (2010).

[22] Shubo Denga**,** Preparation, characterization and application of a Ce–Ti oxide adsorbent for enhanced removal of arsenate from water. Journal of Hazardous Materials 179 (2010).

[23] Eleonora Deschampsa**,** Removal of As(III) and As(V) from water using a natural Fe and Mn enriched sample, Water Research 39 (2005).

[24] http://www.pakissan.com/english/allabout/orchards/citrus/index.shtml

[25] F.A. Cotton, G. Wilkinson, C.A. Murillo, M. Bochmann, Advanced Inorganic Chemistry, 6th ed., John Wiley and Sons, Inc., Singapore, (1999).

[26] H.S. Altundogan, S. Altundogan, F. Tumen, M. Bildik, Arsenic removal from aqueous solutions by adsorption on red mud, Waste Manage. 20 (8) (2000) 761–767.

[27] J. Hlavay, K. Polyak, Determination of surface properties of iron hydroxide coated alumina adsorbent prepared for removal of arsenic from drinking water, J. Colloid Interface Sci. 284 (2005) 71–77

[28] M.C. Teixeira, V.S.T. Ciminelli, Development of a biosorbent for arsenite: structural modeling based on X-ray spectroscopy, Environ. Sci. Technol. 39 (3) (2005) 895–900.

[29] V. Lenoble, C. Laclautre, V. Deluchat, B. Serpaud, J.-C. Bollinger, Arsenic removal by adsorption on iron (III) phosphate, J. Hazad. Mater. 123 (2005) 262–268.

THE CLASSIFICATION AND ANALYSIS OF MANET ROUTING PROTOCOLS

Shahzana Memon*, Pardeep Kumar*, Intesab Hussain Sadhayo*

ABSTRACT

In Mobile Ad-hoc Network (MANETs) mobile nodes are grouped together by using wireless medium that can fail due to its dynamic infrastructure and mobility. Routing protocols are used to facilitate communication among mobile nodes. To provide communication facility among mobile nodes of network routing protocols are used. The main target of routing protocol is to provide path among nodes that should be correct and efficient. Discovery and maintenance of route should have with minimum bandwidth and overhead. This paper investigates and classify Ad hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR), Destination-Sequenced Distance-Vector (DSDV), Temporally Ordered Routing Algorithm (TORA), Multimedia Mobile Wireless network (MMWN), Zone Routing Protocol (ZRP) and Wireless Routing Protocol (WRP).

1. INTRODUCTION

Communication technologies and networking are getting advance now a days. For common activities cell phones and laptops like portable devices are used anywhere by people. Mobile adhoc network is one of the wireless network which is self-configuring and self-location changing. No base stations are used by MANET, every node of network works as router itself. Movements of nodes in MANETs are arbitrarily a change in topology occurs unpredictably. In addition, range Of nodes for transmission are limited in MANET. Due to these limits no direct transmissions can be done by same node [1,2]. Multiple hops are taken for potential path in MANETs, routing is challenging issue in MANET that is complicated due to mobility of nodes. Changes in routes may occur frequently, update of communication links continuously occurs, and sending of messages also done frequently, Hence creation of traffic arises due to this control. Common specification in different devices used by MANET is its limited energy [3].In energy consumption conductive radio waves, resubmission, collision and transmission all are effective. For this purpose strong protocols are needed that manages effective and efficient energy by using different techniques. To achieve efficient routing multiple routing protocols have been proposed. Every algorithm performs different task like, discovering route or maintaining existing known routes. Table driven or proactive routing protocols and on demand or reactive routing protocol are two categories of MANET routing protocol [4].

In this category routing information is stored in routing tables maintain by each node of network. Latest view of Benefit of Table driven protocols is that they do not need to initiate route discovery procedure to reach up to destination but they need massaging overhead because they need to maintain continuously up to date information in its routing table that reduces throughput, consumes power and bandwidth specially in large area networks.

These algorithms create routes when node demands for it. Route discovery procedure is invoked by these protocols during transmission between source and destination. These routes remain active until transmission becomes complete or when they are not used for long time [5]. Benefit of this category protocol is reduced massaging overhead but for new route discovery, delay occurs that is main drawback of these protocols [6].

Energy consumption is an important factor while dealing with MANET, because every node of network contains batteries with some limited power supply for processing. This is challenging issue in MANET because each node forwards packages between nodes as a router and an end system itself so additional energy is needed [1]. In this paper we deal with following metrics to compare different MANET protocols. Routing Overhead

- End to end delay
- Throughput or PDR
- Packet retransmitted
- Energy consumption
- Dropped packets

This paper is organized as follows; in section 2 the working of routing protocols AODV, DSR, DSDV, ZRP, WRP, MMWN, and TORA is given. In section 3, performance metrics are taken in to consideration, while comparison result is represented in section 4 and section 5 contains conclusion of paper.

2. WORKING OF DIFFERENT MANET ROUTING PROTOCOLS:

2.1 AdhocOn Demand Distance Vector:

On the combine features of DSR and DSDV working of

* Department of Computer Systems Engineering,
Quaid-e-Awam University of Engineering Science & Technology, Nawabshah Sindh Pakistan

AODV depends on them. No route maintenance is done by AODV between nodes in the network but demand of route is made by network nodes then routes are discovered and maintained. Unicast route establishment contains some steps those are:

2.1.1 Route Discovery

When any source node S needs to transmit packet to destination node D, it first check in its routing table for route entries towards that destination node. If there is any entry regarding that route is available in routing table then it transmits the packet to the appropriate neighbor towards destination and if it is not there then procedure of discovery of route is initiated by sending RREQ packet from source node. This packet contains source IP address, Destination IP address, its current sequence number, last sequence number and ID of broadcast. Every time when route request is initiated by source node it increments broadcast ID. IP address and Broadcast ID combine makes a unique identifier for RREQ packet and for making timeliness of every data packet sequence numbers are used. When route is source to its neighbors source setup the timer for reply. Every node maintains a reverse path to source in its routing table during the process of RREQ. This will help to forward RREP to source. Lifetime is entered with reverse route entry in its routing table, if route is not used within that life time it will be deleted. AODV also allows initiating route discovery process in case if RREQ is lost during transmission [5,8]. Figure 1 illustrates route discovery procedure in AODV routing protocol.

2.1.2 Expanding Ring Search Technique:

During broadcast of RREQ between source, intermediate nodes and destination, source node uses expanding ring search technique in order to control RREQs network broadcast. (TTL) time to live value is used by this technique, TTL value of RREQ is set by source node as its initial value. If source node does not get any response within that time of discovery, the TTL value is incremented and next RREQ will start broadcasting. TTL incremented value process will continue until it reaches to its threshold value. Across entire network RREQ is broadcasted after reaching to threshold [5,8,9].

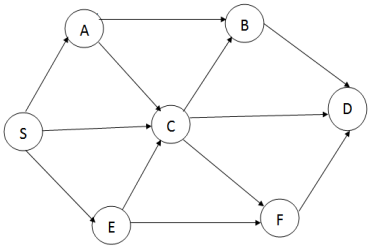


Figure 1: PREQ diagram for AODV when source S sends a request to destination

2.1.3 Setting up of forward path:

After receiving RREQ from source, every intermediate or destination node creates a RREP and sending back to source by unicasting technique. When a reverse path is generated, and RREP is routed back, an entry about that route will be entered in routing table of destination. When source node receives RREP successfully that means route is created between source and destination and transmission of data can be begin by source node [5,8,9].

2.1.4 Route Maintenance:

Due to mobility in adhoc networks route discovery process can be reinitiated by source node in case if source node has taken move from one position of network to another. During the movement of neighbor nodes or destination nodes their upstream nodes will initiate RERR (route error message) and broadcast it to its active predecessor nodes that are affected by that movement. After receiving route error message the source node either reinitiates route discovery process or stops transmission of data depending upon requirement of route.

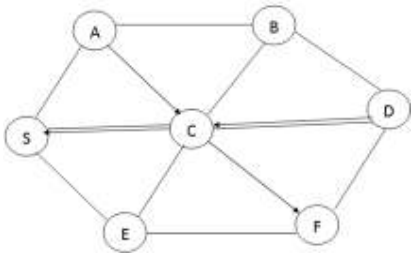


Figure 2: PREP diagram for AODV when destination D sends reply to source S.

Unicast and multicast transmission of packets are supported by AODV protocol. Instead of shortest route, least congested paths are favored by AODV. In case of topological changes AODV gives quick response to its effected active paths. On data packets no additional overheads are put by AODV. Self-detection of medium broadcast between nodes is expected by AODV. Valid route expiration is also possible and finding that expiry time is difficult. Performance of different metrics may decrease as the size of network increase [5,8,9].

2.2 Dynamic Source Routing Protocol:

DSR is an on-demand routing protocol. It creates routes on demand bases instead of hop by hop routing. This algorithm is particularly design for wireless mobile nodes adhoc networks by using multi hop techniques. DSR permits network to be self-configuring and self-organizing and do not need any currently existing topology infrastructure. Route discovery and route maintenance are the two main parts of this protocol. Information about the discovered path/route is stored in cache and that cache is maintain by each node of network.

When source node S needs to transmit packet to destination node D it firstly checks in its cache about the route entry, if that route is available in its cache then it transmits massage through that path and if route to destination is not there and entry of that route is deleted from cache due to remain idle for long time then sender will start broadcasting packet to all its around to ask from neighbor nodes for available route towards destination. Upto the discovery of required route, sender node will wait [5].

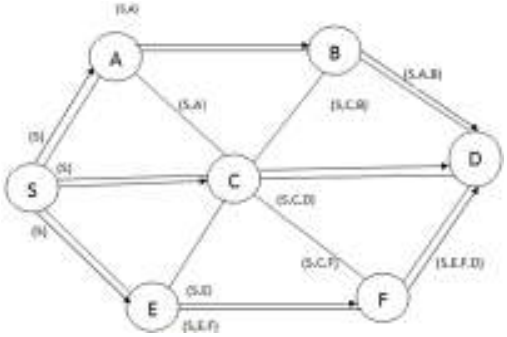


Figure 3: Route discovery diagram for DSR when source S sends a request to destination D.

Till the route is discovered sender node is free to perform other transmissions to different nodes.When neighbor nodes receive route request,they check in its cache whether the required route to destination is available or not. If any node have route information in its cache they send route reply packet o source.They insert an entry about that route in its cache also, so that it can be used in future. When intended packet transmission started and packet is received by any intermediate node then it checks address stored in packet header if it matches then it receives that data and if it does not matches then forwards it further according to route information attached on that packet. Discovery of route procedure is described in Figure 3.Route maintenance is available in DSR algorithm because adhoc networks routes can be failed any time.Route maintenance procedure keeps in view whole network constantly, if there is any route failure then it changes its route cache entries. Figure 4 illustrates route reply procedure between source node and destination node using DSR protocol.

Advantages:

- It does not need to be store routing table because it contain route address in its packet header

Drawbacks:

- Increase in Routing overhead is main issue when topology of network changes or new route needs to be discovered.

By using intelligent caching techniques at network nodes we can reduce overhead but have expensive memory and resources of CPU. Every packet includes source route header that requires remaining bandwidth overhead. The main problem of DSR is scalability, when route queries arise at nodes they use route caches that causes continues repetitive updates and uncontrolled replies at caches on host nodes. The flooded Propagation of all query massages cannot be stop by using earliest queries. As the size of network increases, the size of control packets and massage packets also increases. This causes the degradation in the performance of protocol after sometime [8].

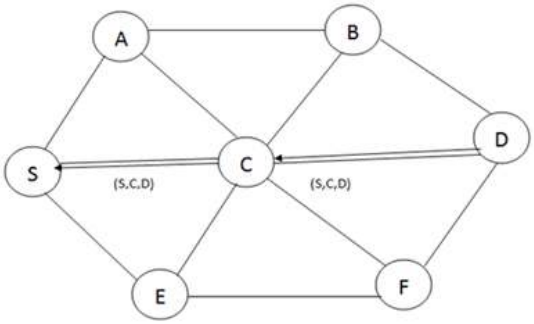


Figure 4: RREP diagram for DSR when source S receives Route reply from Destination D

2.3 Temporary Ordered Routing protocol:

Multihop dynamic networks are operated by using this highly adaptive routing protocol TORA.When source needs to transmit data to any destination the direction of link between them is created by height parameter used by this protocol. Towards a simple destination there may be multiple routes available but it is not important that it contains any short route. When source needs to create route it starts sending QUERY packets to all its around nodes. All networks nodes on receiving that QUERY packet forwards that packet to all its other nodes of network until destination found. Node that receives this QUERY packet, they send UPDATE packet that contains a value of height greater than its neighbor height value of received update packet node. Through this it creates link between original senders of packet to the node that has created that packet.

When node confirms about expire of route towards the destination, so node will set its height and transmits UPDATE packet to its neighbors. If towards destination there is infinite attempt of nodes then route discovery procedure is initialized again as described above. When partition of network occurs then node creates CLEAR packet in order to cancel routing in adhoc network [5,11]. Figure 5 and 6 gives pictorial description for working of TORA routing protocol.

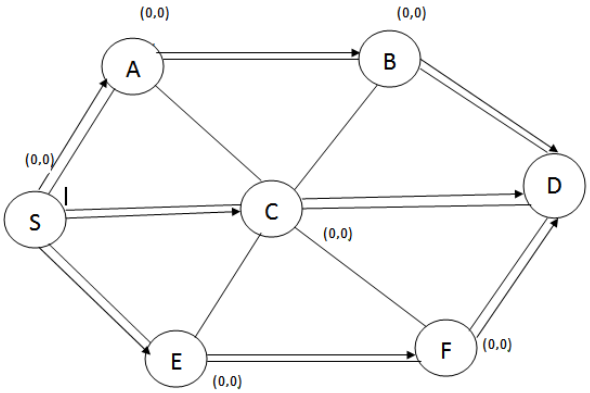


Figure 5: Broadcast Query diagram of TORA when Source S broadcast packet to Destination D

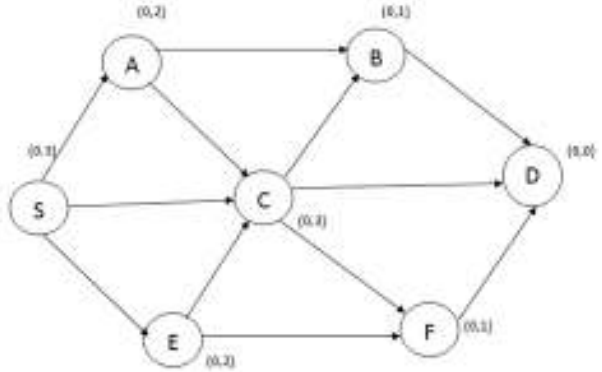


Figure 6: Example diagram for distribute UDP packet in TORA

Advantages:

(i) This protocol supports multiple routes in shape of pairs between source and destination.
(ii) In case of failure or breakage of node switching to an alternate node is done quickly without creating any intervention in source.

Disadvantages:

1 It creates temporary routes between nodes.
2 Nodes contain synchronized clocks among them
3 TORA algorithm depends on those clocks

2.4 Zone Routing Protocol

ZRP is hybrid type of protocol that combines routing techniques of both reactive and proactive algorithms. In ZRP complete network is divided into zones. Each small part of network is represented by zone and each node of zone maintains its routing table that maintains entry information of all other nodes in its zones.Each node can enter inside the network by defines range (in hops). A routing zone is created by each node. ZRP consist three components.

- Intra zone routing protocol: It maintains information about states for links to any given node those are on short distances.
- Reactive inter zone routing protocol: To determine routes that are on some long distance it uses route discovery protocol.
- Border cost resolution protocol: To deliver packets towards nodes that relay on border of zones it uses uni cast routing.

Inside zone maximum number of hops towards farthest node is represented by zone radius d=2 illustrated in figure 7. Within zone table driven architecture is used by nodes and outside nodes of its zone does not maintain routing information record permanently. Nodes that are within the routing zone can send packets immediately because they have immediate routing available but when they need to send packet outside of zone they will get route on demand basis by using any on demand routing protocol. There are three components used by ZRP protocol within table driven or proactive zone, the routing information is managed by (IARP) Intra zone routing protocol. Implementation of IARP is dependent; either you use distance vector routing or link state routing.

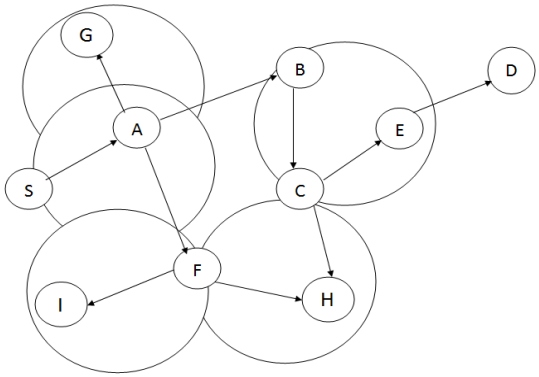


Figure 7: Example of ZRP where S performs route discovery for D with zone radius = 2

(IERP) inter zone routing protocol is used to perform transmissions outside of zone. Within zone routes to nodes are provided by IARP protocol and outside of zone routes are discover by RREQ and RREP packets with the use of IERP protocol [7,8].

Advantages:

(i) When we compare it with table driven protocols, it has reduced amount of overhead during communication and also reduces delays.
(ii) Routes was discovered faster due to association of DSR protocols because to travel outside the zone it needs route and the route will be discover by routing up to boundaries of required destination.
(iii) To the destination routes are maintain by boundary nodes using proactive techniquesnodes of

boundaries sends reply back to the source to create route between source and destination by using address of routing.

Drawbacks:

1 Protocol performs transmission using proactive protocols when large values are needed to be used for routing zone, while for smaller values it perform transmissions like reactive protocol.

2 Proactive overhead is limited by ZRP to the zone size only and reactive overhead is also limited by ZRP to choose border nodes only.

3 When through the whole network RREQ packets are started flooding theninefficiency may arise. Better solution is provided by this protocol to some extent in order to reduce overhead and delay during communication but this is beneficial only for zone size and dynamics of zone.

4 When through IERP destination has found then optimized shortest path does not provided by ZRP. As size of network increases, large overheads are created by ZRP periodically. This means every node needs large memory and topological information at high level and extra resources are taken b network as a burden.

2.5 Multimedia support in Mobile Wireless networks:

This protocol maintains the network by using hierarchy of clusters. Two nodes are used by each cluster end points and switches. Location management is performing by Location manager (LM) that each cluster contains. Multimedia support in Mobile Wireless network (MMWN) contains a database that is dynamically distributed and stores all information in it. For user data traffic sources and destination nodes are only can be an end point and routing functions can be perform by switches [7]. When formation of low level partitions are needed in hierarchy, most convenient switches are choose by end points for checking their associatively with other nodes. Cells are organized by grouping end points around those switches called cluster heads. This process is known as "cell information". Clusters are formed hierarchically by switches and each of cluster functions as multi hop packet radio network. Keeping track of hierarchical addresses is important to support transfer of data between mobile nodes. When changes in hierarchical addresses occur due to mobility, location manager uses both paging and query/response in conjunction. Every cluster contains its own LM that controls all the nodes within network and outside of cluster. With respect to clustering hierarchy each node has a roaming level. Within roaming cluster paging technique is used to locate mobile nodes. When mobility of nodes occurs and node moves outside of current roaming cluster, update about location is transferred to LM [16].

This scheme contains several features whose implementation is complex.

i. Location manager is connected with hierarchical topology of network. Location finding and location updating becomes complex due to this feature. Through Location managers hierarchical tree, location updating/finding has to travel. Due to changing in location managers hierarchical cluster membership, reconstruction in location management tree occurs.

ii. For routing and addressing MMWN provides its own routing protocols. It is not based on IP standard protocol. To enable internetwork of MMWN encapsulated interface translation capabilities is needed with IP networks [16].

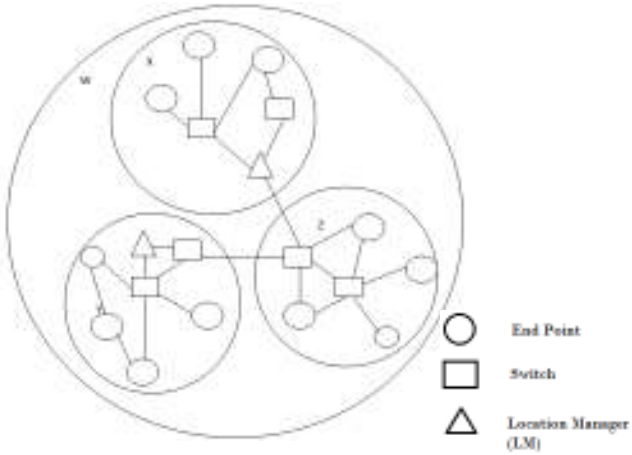


Figure 8: Example of Clustering Hierarchy in MMWN

Benefits are when we compare this algorithm with other traditional table driven architectures it reduces routing overhead due to use of LM because LM performs all functions of location finding and location updating. Location finding and updating functions are very difficult due to networks hierarchical structure and LM is very near to it. In order to perform location finding and updating it needs to travel massages through LMs hierarchical tree and membership of LMs in hierarchical tree changes makes consistency management very difficult and also effects hierarchical tree of management. Problems are created due to that are difficult to solve [7].

2.6 Destination Sequenced Distance Vector:

DSDV maintains view of the network continuously or DSDV keeps in view whole network consistently by maintaining routing tables. Routing tables contain routing information received by periodic update of routing. Every node contains routing table storing routing information in it. New route broadcast contain destination address, no: of hops required to reach destination, destination sequence no: & for new broadcast a new sequence no:. A fresh

route is that receives currently new sequence no: and sequence no: of two routes are suppose to be same than by considering better metric new route will be selected. DSDV uses distance vector shortest path routing algorithm that creates a single link towards the destination. It uses two type of packet updates in order to minimize amount of overhead. First one is full dump packet to carry routing information that is available. Second is incremental packet that carries current information only (changed since last full dump) as compare to full dump packets more amount of incremental packets are sent.

Regardless of network traffic DSDV needs to transmit update packets about routing table periodically so it still has large amount of overhead and according to O(N2) overhead grows. Due to scalability of network more bandwidth and more size of routing tables are required. DSDV is not efficient in large networks because when network topology change occurs routing loops can occur.

In case of link failure and addition of nodes setting time is complex to determine. Multipath routing is not supported by DSDV. In terms of routing overhead DSDV is good to perform tested by different simulation environments but because it cannot detect link breakages fastly so more packets of data are dropped. In case of security issue DSDV specifications are silent. DSDV trust all nodes and assume them as cooperative but once attacker creates a false sequence, it sends new packets consistently to update the value so huge part of network nodes will be cheated and du to misbehaving of single host node entire network can be effect by serious threats [7,8]. Figure 9 illustrates the working of DSDV protocol.

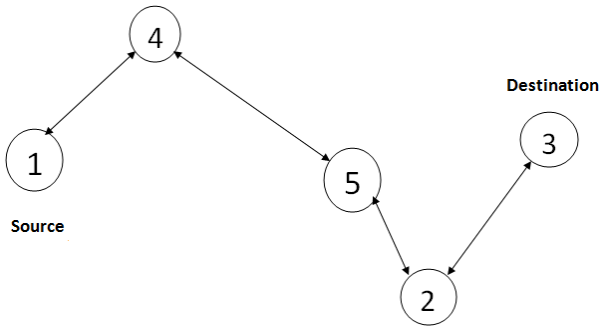


Figure 9: Example of DSDV

In DSDV each node of network maintains a routing table that contains addressing information of every other node in the network. In order to reach destination node each node of network contains address of next hop node along with its address in routing table. Figure shows bi directional connection of DSDV network where node 1 is source and node 3 is destination. Node 1 transmits packet for further forwarding to node 4. Node 4 will look for address of destination in its routing table. Node 4 then takes next hop to transmit packet towards node 5 in this case. This procedure is repeated consistently until packet reaches to its required destination.

2.7 Wireless Routing Protocol:

Same as DSDV all nodes maintain routing information among them. Four tables are maintaining by each node namely Distance table, Routing table, Link cost table and Message retransmission list (MRL) table. MRL contains update massage sequence number, a transmission counter, a flag vector required by acknowledgment (by vector neighbor entry) and update list of every update massage in its each entry. MRL takes all information about massages that are needed to be transmitted and transmission acknowledgment given by which neighbors is recorded in update massages.

WRP needs huge amount of memory to maintain four tables as compare to other proactive routing protocols. It creates link between nodes by sending hello messages. Hello messages are broadcasted when network nodes are free from transmission so each node needs to remain active every time. Nodes cannot enter in idle condition to save their battery power because it maintains separate tables so it has less latency as compare to others. Underlying routing protocols used by WRP is distance vector shortest path routing and their addition and link failure, it is complex to manage. During broadcasting message WRP focuses on near vicinity and even in their vicinity neighbor nodes may not have complete information is concluded by WRP so there is a limit of transmission of data in small regions. To neighbor nodes they have limited update messages. This limit is not for close vicinity but for network view of nodes [8].

3. PERFORMANCE METRICS:

Performance of adhoc routing protocols is compare by taking following qualitative and quantitative metrics in consideration.

(i) Routing Overhead: To Broadcast/propagate data packets for route discovery and route maintenance how many packets need to be sent. Due to location changing in network by nodes, unnecessary routing overhead occurs because in routing table the generation of stale routes occurs.

(ii) Throughput: it is the ratio between amounts of total packets sends by sender to receiver and what amount of time receiver takes to receive last packet. It is measured in Bits per second (bps). Throughput is one of the main parameter of MANETs network that provides capacity of channel that is use full for transmission. It performs selection of destination node at the start of simulation and gives information about data is delivered to destination correctly or not.

(iii) Packet Delivery Ratio: it is the ratio between total amount of packets incoming through channel and data packets those are received successfully.Loss rate will be described and seen at transport protocols that effects network supported maximum delay. When load of network and no of packets are less, DSR performs very well. Performance of DSR is effected by increase of nodes in network in other words due to increased network traffic. AODV performance remains same continuously while DSDV performs better as compare to other two protocols.

(iv) End to end delay: it is defined as how much time packet takes to be transmitted from source to destinations application layer. It is measured in seconds. Delays those occur due to queue in transmission of data packets and route discovery process are also included in end to end delay.CBR packets high rate effects delay. Buffers become more faster because before transmission packets need to be stored in buffers for long time. This is seen during research that AODV and DSR performance in terms of average end to end delay is same almost. But due to increase in no: of nodes routing tables load exchange also increases so the DSDV performance degrades and increase in frequency of exchange also occurs due to nodes mobility.

(v) Energy variance of node: calculation of total amount of distributed energy among nodes.

(vi) Remaining battery power: this metric calculates remaining average battery power remaining and total no: of nodes in the network. In adhoc network every node is operated on power of battery and rare resource is energy of battery. Performance of protocols is analyzed in terms of power through this metric.

(vii) Consumed power: this metric considers average of consumed batter power v/s no: of nodes in network.Energy consumption also occurs during hearing process not only during transmission and reception of data.

(viii) **Retransmission:** no of packets that are retransmitted because of packet loss. when sender sends data packet to destination it retains a copy of packet until it receives acknowledgment from receiver that it has receives the packet correctly. Sender performs retransmission in different circumferences. Reasons for retransmission are:

1 If time exceeds and no acknowledgment found at sender.
2 When sender observes about unsuccessful transmission.
3 When sender gets notice from receiver that expected data is not received.
4 When data packet is received by receiver but not correctly then notifies to sender about correctness.

(ix) **No: of dropped packets:** amount of packets that are failed to be reach at destination during transmission is known as packet drop. This metric calculates no: of dropped packets vs. time. Due to completion of TTL (time to live) dropping of packets occurs. Life time of packets depends upon protocol. Packets became dropped when protocol takes too much time to decide path towards destination. Routing directions can be find by efficient protocols then rate of dropped packets reduces. The dropped packets for DSR are less than that of DSDV.

4. DISCUSSION / COMPARATIVE STUDY

Our main objectives were minimizing network throughput, maximizing lifetime of network and minimizing delay during transmission and hearing for channel. This paper evaluated the comparative study and performance of DSR, AODV, TORA, DSDV, WRP and MMWN on basis of energy efficiency, throughput, routing overhead, end to end delay and no: of packets dropped. We have concluded that inefficient performance was given by TORA protocol. In small networks, no any significant differences reveal by throughput and energy consumption. In small network sizes DSR,DSDV and AODV performance was comparable but in large networks good results are only produce by AODV and DSR. While in all scenarios in terms of throughput AODV performance was overall good. In terms of energy consumption performance of AODV and DSR was comparable while TORA, DSDV and ZRP consume more power but dropped packet ratio was maximum in AODV as compare to other routing protocols.

The following table illustrates the routing protocols comparison in terms metrics discussed in section 3.

5. CONCLUSION

In this paper, we have described working of several proposed routing schemes. According to routing strategy, we have classified these schemes on basis of table driven, On–demand and hybrid type of routing algorithms. We have compared these three categories of routing algorithms highlighted their characteristics, differences, features, benefits and their drawbacks. Mobile networking (MANET) field is expandable and changes day by day. Still some of the routing protocols working management and problem are too much complex to solve as to write new protocol. Bandwidth constraints and limited power of mobile devices are most prominent issues in MANET. Security and power awareness is hard to achieve due to dynamically changing topology. These issues are handled by some schemes that are mentioned above. Therefore a better routing solution is needed that also address other issues related to routing. Our future work is the focus on the study of these issues and for MANET routing protocols we will take effort to propose the solution for power awareness and secure routing.

Routing Protocols	Performance Metrics										
		Routing Overhead	End to End Delay	Throughput	Packet Retransmitted	Energy Consumption					Dropped Packet
						Consumed Power	Remaining Battery	Tx	Rx	Routing	
	AODV	High	Medium	Medium	High	Low	High	M	H	M	High
	DSR	Medium	High	Low	Low	Low	High	M	H	L	Low
	TORA	Medium	H.L	L,H	High	High	Low	H	H	H	_
	DSDV	High	Low	Medium	_	Stable	Low	M	H	M	High
	ZRP	_	Medium	Low	_	High	Low	H	H	H	Medium

Table 1: Comparison of the routing protocols with respect to the performance metrics

REFERENCES

[1] A.A.A. Radwan ,T.M.Mahmoud. E.H Houssein. Evaluation comparison of some MANET networks routing protocols. Cairo University Egyptian Informatics Journal, vol. 12, 95-106, 2011.

[2] Tyagi SS, Chauhan RK. Performance analysis of proactive and reactive routing protocols for MANET networks. Int J Comput. Appl., vol. 1, No. 4, 2010.

[3] S. A. Ade, P.A.Tijare. Performance Comparison of AODV, DSDV, OLSR and DSR Routing Protocols in Mobile MANET Networks. International Journal of Information Technology and Knowledge Management, Volume 2, No. 2, pp. 545-548.

[4] Khaleel Ur Rahman Khan Rafi U Zaman A. Venugopal Reddy, Performance Comparison of on-Demand and Table Driven Ad hoc Routing Protocols using NCTUns, Tenth International Conference on Computer Modeling and Simulation, Pages 336-341, 2008.

[5] Sunil Taneja and Ashwani Kush. A survey of Routing Protocols in Mobile Ad hoc Networks. International Journal of Innovation, Management and Technology, Vol. 1, No.3. 2010.

[6] Kanakaris, Venetis, Ndzi, David and Azzi, Djamel. Ad-hoc networks energy consumption: a review of the ad hoc routing protocols. Journal of engineering Science and Technology Review (JESTR), vol. 3, No. 1, pp.162-167, 2010.

[7] Mehran Abolhasan, TedeuszWysicki, ErykDutkiewicz. A review of routing protocols for mobile ad hoc networks. Ad hoc networks, vol. 2, No. 1, pp. 1-22, 2004.

[8] B. Bellur, R.G. Ogier, F.L Templin. Topology broadcast based on reverse path forwarding routing protocol, IEEE Eighteenth Annual Joint Conference of the Computer and Communications Societies, vol. 1, 1999.

[9] S.Das, C. Perkins, E. Royer. Ad hoc On demand Distance vector (AODV) routing, Internet Draft, 2003.

[10] HumayunBakht. Survey of Routing Protocols for Mobile Ad hoc Networks. International Journal of Information and Communication Technology research, vol. 1, No. 6. 2011.

[11] MamtaDhanda, ShikhaChaudhry, Survey of Routing Protocols for Mobile Ad Hoc Networks, International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 4, 2013.

[12] P.Kuppusamy, Dr.K. Thrunavukkarasu. Dr.B.Kalaavathi, A Study and Comparison of OLSR, AODV and TORA RRouting Protocols in Ad Hoc Networks, 3rd International Conference on Electronics Computer Technology (ICECT), vol. 5, pp. 143-147, 2011.

[13] Ali Norouzi and A. HalimZaim. Energy Consumption Analysis of Routing Protocols in Mobile Ad Hoc Networks, Real-Time Systems, Architecture, Scheduling, and Application, Dr. Seyed Morteza Babamir (Ed.), 2012.

[14] Nurul I Sarkar. MANET Routing Protocol Comparison: Joint Node Density and Mobility, Journal of Communication and Computer, vol. 8, pp. 1033-1038, 2011.

[15] S. Sathish, K. Thangavel and S. Boopathi. Performance Analysis of DSR, AODV, FSR and ZRP Routing Protocols in MANET, Performance Analysis of DSR, AODV, FSR and ZRP Routing Protocols in MANET pp 57 – 61

[16] Guangyu Pei and Mario Gerla, Mobility Management in Hierarchical Multi-hop Mobile Wireless Networks, ICCCN, pp. 324-329. 1999.

ANALYSIS OF AD-HOC ON-DEMAND DISTANCE VECTOR ROUTING PROTOCOL AGAINST NODE MISBEHAVIOR ATTACKS IN WIRELESS SENSOR NETWORKS

Adnan Ahmed*, Umair Ali Khan*, Adnan Rajpar**

ABSTRACT

Wireless Sensor Networks (WSN) have gained remarkable appreciations over the last few years. Despite significant advantages and tremendous applications, WSN is vulnerable to variety of attacks. Due to unattended nature of WSN, sensor nodes are more prone to be overtaken by an adversary. By doing that, an adversary can learn the contents of the victim's memory, can have access to valid cryptographic keys, and can also modify the behavior of corrupted nodes. In this paper, we investigate some of the most severe node misbehavior attacks in WSN, namely blackhole and grayhole attacks, using Ad-hoc On Demand Distance Vector (AODV) routing protocol. A detailed NS2 based implementation and comparative analysis of these attacks has been presented. The performance of AODV is evaluated by considering different metrics such as packet delivery ratio, packet drop ratio, average end-to-end delay, normalized routing load, and energy consumption. Simulation results are provided to show the effects of these attacks on AODV protocol which suffers from increased packet loss and decreased delivery ratio. Some counter measures against node misbehavior attacks are also provided.

Keywords*: AODV; grayhole attack; wireless sensor networks; security; node misbehavior; blackhole attack.*

1. INTRODUCTION

Wireless Sensor Network is a self-organized network comprising of individual nodes that connect with their surroundings by sensing or controlling physical parameter. A wireless sensor node is equipped with micro-sensor technology that has low computational power, low signal processing power, limited energy resources and short-rage communication facility [1]. Recent advances in computing and communication have enabled wireless sensor network to be deployed in variety of applications such as battle field monitoring, battle damage assessment, environmental monitoring, smart environments, monitoring the status of structures such as bridges, factory process control and automation, vehicle tracking and detection and monitoring disaster area [2].

Sensor nodes are placed in large number in hostile environments, which makes it difficult to protect them against tampering or captured by an adversary force that can launch attacks to make a node compromised and can have easy access to valid keys and memory contents [3]. This unattended nature of WSN makes sensor nodes vulnerable to node physical capture, selfish and malicious behavior of nodes. Routing in WSN is a cooperative process where routing information must be shared between all the nodes on the route to destination. There might be a strong case that some malicious, selfish or misbehaving nodes might exist on a discovered route and may not fulfill the desired rules and regulations of the protocol. In this study, we will analyze the impact of two types of node misbehavior attacks: blackhole and grayhole.

Prior to proposing a secure solution to protect WSN against the aforementioned node misbehavior attacks; it is important to gain full understanding of how these attacks are launched. The major objective of this paper is to analyze how these attacking nodes exploit the weakness of a route discovery mechanism of a routing protocol and work maliciously. Most of the existing studies on performance evaluation of WSN mainly focus on investigating the impact of attacks without providing solutions to avoid them. WSN is a resource constraint network and energy is most critical design parameter for providing secure solutions, but most of the literature did not pay much attention to analyze the impact of attacks on the overall energy consumption of the network. In this study, along with providing countermeasures against node misbehavior attacks, impact of the attacks is also investigated and analyzed with justified parameters for WSN.

The rest of the paper is organized as follows. Section 2 provides the overview of blackhole attack and grayhole attack; and also discusses the related work in this domain. Section 3 presents the simulation model for simulating attacks. Section 4 presents the simulation results with analysis. Section 5 provides some countermeasures against node misbehavior attacks. Section 6 concludes the paper with some potential future work.

* Department of Computer Systems Engineering,
** Department of Information Technology,
Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Pakistan.

2. BLACKHOLE AND GRAYHOLE ATTACK MECHANISM

In this paper, blackhole and grayhole attacks are considered as an attack model. In a blackhole attack [4–6], a compromised node claims itself to be the most suitable forwarding node but refuses to cooperate with routing rules and drops all the received packets. A malicious node advertises itself to have fresh and optimal path to the destination by exploiting weaknesses of the route discovery packets (RREQ and RREP). A blackhole node sends false RREP packets to attract most of the network traffic by incorporating highest sequence number in RREP packets. Figure 1 shows the behavior of blackhole node in a network using AODV routing protocol.

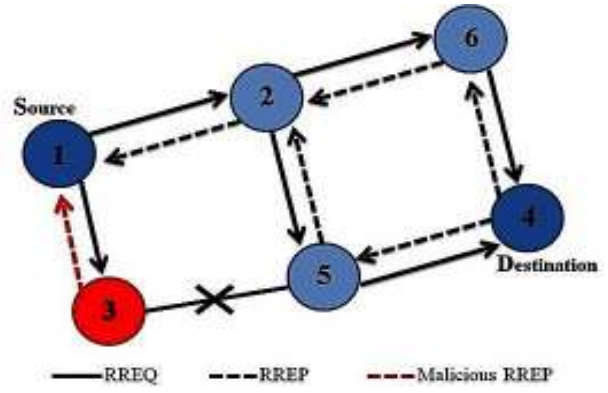


Figure 1: Blackhole attack

As shown in Figure 1, when a source node wants to send packets to destination node, it broadcasts RREQ packets. The node receiving RREQ packets responds with RREP packets. The malicious node (node3) sends a false RREP packet with highest sequence number indicating it has better route to destination. The source node assumes that the provided information is true. Therefore, it sends packets to the malicious node. The malicious node exhibiting blackhole behavior drops all the received packets, leaving none or very few packets to reach the destination. The most critical influence of this attack on the network results in severely diminishing the packet delivery ratio.

A compromised node against grayhole attack drops packets selectively rather than dropping all received packets [7–9]. The selective dropping depends on the type of packets or group of some nodes. For example, a grayhole attacker node forwards UDP packets while dropping TCP packets. Another malicious behavior of the very attack is to drop packets for particular time duration and at later time switching its behavior to normal node. The grayhole node sends genuine route-reply packets in contrast to blackhole nodes where they send fake route-reply to attract most of network traffic. Due to the aforementioned behaviors of grayhole node, detection becomes very difficult [10][11].

Let us assume that the grayhole node is represented by M, with aim to drop packets for some time period. Let $P(M)$ denotes the probability for dropping packets for grayhole node and $P(N)$ denotes the probability of normal nodes. The probability for the occurrence of grayhole attack in WSN is given by equation 1.

$$P(M) = \frac{P\left(\frac{M}{N}\right) * P(N)}{P(M)} \qquad (1)$$

3. RELATED WORK

The relevant literature shows a number of studies investigating the performance of WSN and MANET under node misbehavior attacks. The authors in [12–14] provide theoretical analysis of various node misbehavior attacks, but none of the attacks is simulated on either of proactive or reactive protocols to study the effects of the attacks.

In [15–17], the performance of MANET in presence of wormhole attack is analyzed. In wormhole attack, an adversary creates a connection (called tunnel) between two different points in the network that are not in the communication range of each other. The two colluding nodes under wormhole attack capture packets at one end (source) and tunnel them to other end (destination) and replay them. The authors in [15] implement Packet Leash and Time of Flight techniques to detect and prevent a wormhole attack. The authors in [15][16] do not provide any simulation based study to consider the effects of wormhole attack on AODV. The authors in [17] analyze the performance of AODV under a wormhole attack only in terms of throughput with limited network parameters which is not sufficient to measure performance of a MANET.

The authors in [10] analyze the performance of LEACH protocol against a grayhole attack. LEACH protocol is designed for resource constrained WSN where energy consumption is a critical factor. However, the impact of the attack on overall energy consumption of nodes is not given consideration in this study. The relevant literature provides various other studies where impact of attacks on routing protocol has been investigated [16][18–22]. Most of the studies mainly target generic ad-hoc networks which provide powerful hardware platform with enough storage, energy, memory and processing resources but the dynamics of WSN are different where sensor nodes with limited resources are deployed in the network. Furthermore, most of the studies do not pay attention on analyzing the performance in terms of energy consumption and do not provide appropriate measures to

defend those attacks. In order to propose an optimal solution for WSN, the impact of attacks must be analyzed under resource-constrained environment of WSN.

4. THE SIMULATION MODEL

Network Simulator-2 (NS-2), an event based simulator, has achieved tremendous appreciations in network and communication research community due to its capability of analyzing the dynamic nature of networks, its flexible design and modular nature [23]. In this study, performance of AODV and Compromised-AODV (C-AODV) protocols has been analyzed under blackhole and grayhole attacks. The performance of AODV routing protocol is evaluated under these attacks by considering different performance parameters such as packet delivery ratio, packet drop ratio, end-to-end delay, normalized routing load and overall energy consumption of sensor nodes.

Our evaluations are based on the simulation of variable number of malicious nodes (blackhole and grayhole nodes). Fifty sensor nodes are randomly placed to form a wireless senor network over an area of 1000m × 500 m. There are 4 source nodes placed at different locations which transmit packets at specified time period, and one sink node that is placed in the center of network topology. Table 1 lists the parameter settings for our simulation environment.

Table 1: The simulation parameters

Simulation parameters	Values
Simulation Area	1000 x 500 m^2
Simulation Time	1000 sec
Routing protocol	AODV, C-AODV
Number of sensor nodes	50
Number of Malicious nodes	0, 1, 2, 3, 4
Transport layer protocol	UDP
Initial Node Energy	50 joules
Packet size	50 bytes
Node Mobility	Random
MAC protocol	IEEE 802.15.4
Application layer traffic	CBR

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, we analyze how AODV behaves under various number of blackhole and grayhole nodes. Figure 2 shows that when neither of the attack is launched on AODV, Packet Delivery Ratio (PDR) is 97%. When the number of malicious nodes increases in the network, PDR deteriorates. For a blackhole attack, it drops to 70%, 58%, 24% and 2% when there are 1, 2, 3 and 4 blackhole nodes in the network, respectively. For grayhole attack, PDR drops to 74%, 60%, 55% and 42% when there are 1, 2, 3, and 4 grayhole nodes in the network, respectively. Both attacks bring delivery ratio to unacceptable ranges but blackhole attack proves to be more packet-hungry. A compromised node under a blackhole attack sends fake RREP packets informing other nodes that it has the shortest route to destination while dropping all the received packets, leaving none or few packets to reach destination. A grayhole node, on the other hand, sends genuine RREP packets but switches its behavior from normal to malicious or vice versa.

Figure 3 shows average end-to-end delay comparison for AODV and C-AODV. Under normal scenario, packets reach destination within minimum delay. As the density of misbehaving nodes increases, average end-to-end delay also increases. The average end-to-end delay increases by 45% and 80% for grayhole and blackhole attacks respectively, in case where 4 misbehaving nodes are part of the network. Such increased delay caused by both grayhole and blackhole attacks are not acceptable for mission critical applications.

Figure 4 shows the number of dropped packets in the network having some malicious nodes. It is observed that as the density of malicious nodes increases, the number of packet drop also increases, as it is inversely proportional to delivery ratio. Due to the inherent characteristic of high packet drop by blackhole node, the number of dropped packets by a blackhole node is higher than a grayhole node for all cases that probabilistically drop packets leaving a few or more packets reaching the destination.

Figure 5 shows how the normalized routing overload is affected in the presence of misbehaving nodes. As number of misbehaving nodes increases in the network, Normalized Routing Load (NRL) also increases. NRL refers to the ratio of total number of transmitted control packets to the total number of received data packets.

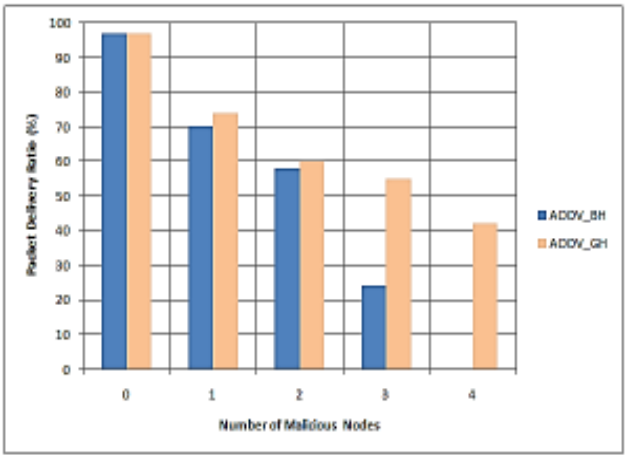


Figure 2: Number of malicious nodes vs. packet delivery ratio

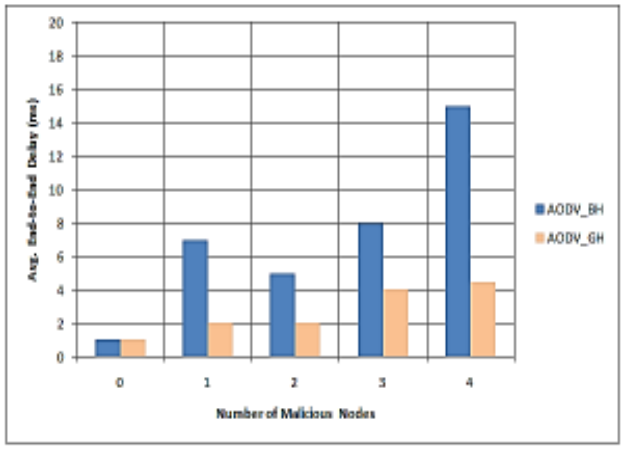


Fig. 3: Number of malicious nodes vs. avg. End-to-end delay

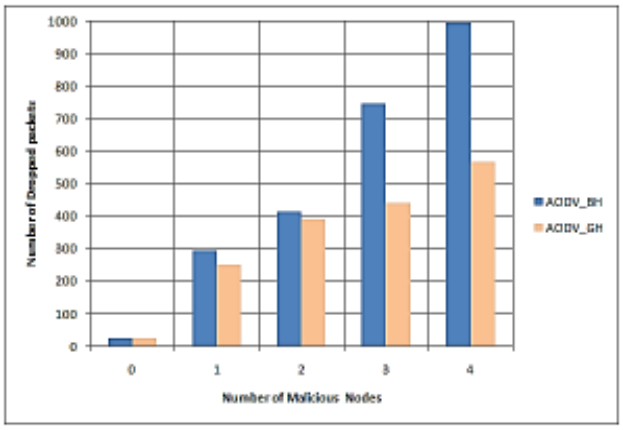


Fig. 4: Number of malicious nodes vs. number of dropped packets

A blackhole node sends false RREP packets and generates additional routing packets which leads to high routing load as compared to a grayhole node. As WSN is a resource-constrained network, such increased overload may badly affect the lifetime of the network.

Figure 6 shows comparison of average energy consumption for both AODV and C-AODV. The energy consumption is directly related to the number of transmitted and received messages (data or control). Simulation results show that as the number of misbehaving nodes increases, average energy consumption also increases due to the adverse effects of the misbehaving nodes on route discovery and route resolve mechanisms of AODV. The increased number of route maintenance calls and route discovery control packets increases overall energy consumption of the network, when it comes under a number of compromised nodes. Such increased energy consumption is not feasible for battery powered sensor nodes which mostly operate in unattended environments.

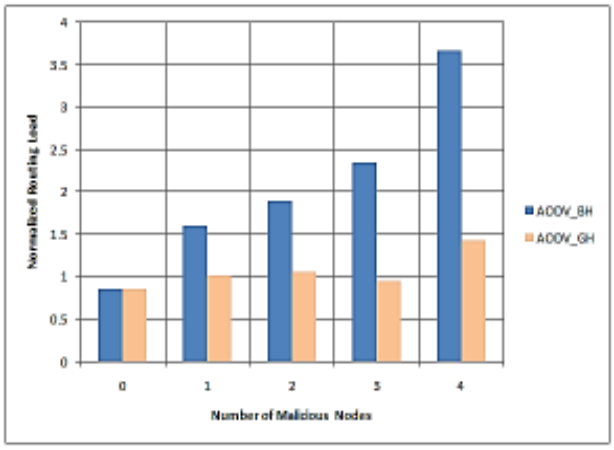


Fig. 5: Number of malicious nodes vs. normalized routing load

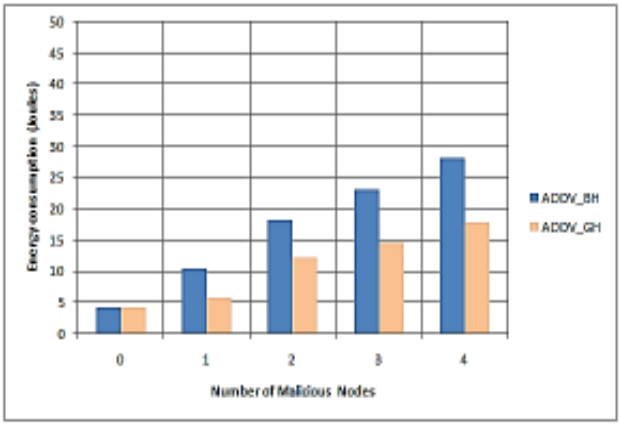


Fig. 6: Number of malicious nodes vs. energy consumption

6. COUNTERMEASURES AGAINST NODE MISBEHAVING ATTACKS

Multi-hop communication requires the exchange of routing information among intermediate nodes. However, multi-hop communication also raises the problem to securely route packets as some misbehaving nodes may become part of the active route to destination. Nonetheless, following defense mechanisms can be adopted for a secure routing.

i. Verifying packet sequence number can be helpful, in few cases, for the detection of misbehaving nodes. A node is considered as a misbehaving node if an abnormal increase in the sequence number is identified.

ii. Secure routing protocols may exploit some mechanisms for providing rewards and punishments based on the behavior of nodes. If a node cooperates in packet forwarding, it may be provided rewards, otherwise punished.

iii. Authentication methods can be used to determine whether the sensor node can participate in routing or

not. SEAD [24] and Ariadne [25] are the two secure routing protocols based on an authentication mechanism which prevent misbehaving nodes to become part of the network.

iv. Game theory based approaches [26] are also useful in dealing with misbehavior nodes. These approaches assume that some greedy actions are performed by malicious nodes to gain better performance, such as leveraging the operating point, "Nash Equilibrium" and higher share of bandwidth.

v. Intrusion detection [27][28] and watchdog [29][30] solutions may be used for monitoring the behavior of nodes. If some malicious behavior is observed, an appropriate action may be triggered like alerting neighboring nodes.

vi. Trust and Reputation based systems [31][32] may be used for detection and isolation of malicious nodes. These systems facilitate the nodes to predict the behavior of other nodes and provide secure mutual interaction.

vii. Exploiting the multi-path routing approach [33] may minimize the adverse effects of misbehaving nodes due to the availability of backup paths. However, this approach is only responsible for minimizing the impact, but does not completely prevent the attack.

7. CONCLUSION

In this study, we comprehensively analyzed the performance of WSN against the most severe node misbehaving attacks such as blackhole and grayhole attacks. NS2 simulator has been used to simulate these attacks using AODV routing protocol. The performance is evaluated in terms of packet delivery ratio, number of dropped packets, average end-to-end delay, average throughput, normalized routing load and energy consumption. Simulation results show how badly these attacks affect the overall performance of a WSN. As the number of misbehaving nodes increases in the network, it badly affect the overall performance of the network and brings the delivery ratio to an unacceptable range. Some countermeasures against node misbehavior attacks are also provided. This analysis is helpful to gain insight into the unexpected anomalies in a WSN and envisaging appropriate counter-measures. Our future work in this direction will focus on implementing other node misbehavior attacks, such as sinkhole attack, Sybil attack, HELLO flood attack, selfishness and wormhole attack in WSN and providing efficient and trust-aware routing mechanisms to counter such attacks.

REFERENCES

[1] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," Ad Hoc Networks, vol. 3, no. 3, pp. 325–349, May 2005.

[2] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," Computer Networks, vol. 51, no. 4, pp. 921–960, Mar. 2007.

[3] M. Momani and S. Challa, "Survey of Trust Models in Different Network Domains," International Journal of Ad hoc, Sensor & Ubiquitous Computing, vol. 1, no. 3, pp. 1–19, 2010.

[4] M. Roopak and B. Reddy, "Blackhole Attack Implementation in AODV Routing Protocol," International Journal of Scientific & Engineering Research, vol. 4, no. 5, pp. 402–406, 2013.

[5] M. O. Pervaiz, M. Cardei, and J. Wu, "Routing security in ad hoc wireless networks," in Network Security, S. C.-H. Huang, D. MacCallum, and D.-Z. Du, Eds. Boston, MA: Springer US, 2010, pp. 117–142.

[6] M. Al-Shurman, S.-M. Yoo, and S. Park, "Black hole attack in mobile Ad Hoc networks," in Proceedings of the 42nd annual Southeast regional conference on - ACM-SE 42, 2004, p. 96.

[7] S. Chundong, P. Yi, J. Wang, and H. Yang, "Intrusion Detection for Black Hole and Gray Hole in MANETs," KSII Transactions on Internet and Information Systems, vol. 7, no. 7, pp. 1721–1736, 2013.

[8] M. Arya and Y. K. Jain, "Grayhole Attack and Prevention in Mobile Adhoc Network," International Journal of Computer Applications, vol. 27, no. 10, pp. 21–26, 2011.

[9] S. Ramachandran and V. Shanmugam, "Performance Comparison Of Routing Attacks In Manet And Wsn," International Journal of Ad hoc, Sensor & Ubiquitous Computing, vol. 3, no. 4, pp. 41–52, 2012.

[10] A. P. Renold, R. Poongothai, and R. Parthasarathy, "Performance analysis of LEACH with gray hole attack in Wireless Sensor Networks," in International Conference on Computer Communication and Informatics, 2012, pp. 1–4.

[11] W. Zada Khan, Y. Xiang, M. Y Aalsalem, and Q. Arshad, "The Selective Forwarding Attack in Sensor Networks: Detections and Countermeasures," International Journal of Wireless and Microwave Technologies (IJWMT), vol. 2, no. 2, pp. 33–44, Apr. 2012.

[12] K. Xing, S. S. R. Srinivasan, M. Rivera, J. Li, and X. Cheng, "Attacks and Countermeasures in Sensor Networks : A Survey," in Network Security, 2010, pp. 251–272.

[13] S. Mohammadi, R. E. Atani, and H. Jadidoleslamy, "A Comparison of Link Layer Attacks on Wireless Sensor Networks," Journal of Information Security, vol. 02, no. 02, pp. 69–84, 2011.

[14] J. young Kim, R. D. Caytiles, and K. J. Kim, "A Review of the Vulnerabilities and Attacks for Wireless Sensor Networks," Journal of Security Engineering, vol. 9, no. 3, pp. 241–250, 2012.

[15] A. A. Bhosle, T. P. Thosar, and S. Mehatre, "Black-Hole and Wormhole Attack in Routing Protocol AODV in MANET," International Journal of Computer Science, Engineering and Applications (IJCSEA), vol. 2, no. 1, pp. 45–54, 2012.

[16] R. H. Jhaveri, A. D. Patel, J. D. Parmar, and Bh. I. Shah, "MANET Routing Protocols and Wormhole Attack against AODV," International Journal of Computer Science and Network Security (IJCSNS), vol. 10, no. 4, pp. 12–18, 2010.

[17] A. Ratmele and R. Dhakad, "Performance Analysis of AODV under Worm Hole Attack through Use of NS2 Simulator," International Journal of Engineering Research and Applications (IJERA), vol. 3, no. 3, pp. 201–205, 2013.

[18] I. Hababeh, I. Khalil, A. Khreishah, and S. Bataineh, "Performance Evaluation of Wormhole Security Approaches for Ad-hoc Networks," Journal of Computer Science, vol. 9, no. 12, pp. 1626–1637, 2013.

[19] H. Simaremare and R. F. Sari, "Performance Evaluation of AODV variants on DDOS , Blackhole and Malicious Attacks," International Journal of Computer Science and Netowrk Security (IJCSNS), vol. 11, no. 6, pp. 277–287, 2011.

[20] K. Pavani and D. Avula, "Performance of Mobile Adhoc Networks in Presence of Attacks," in International Conference on Information Security and Artificial Intelligence (ISAI 2012), 2012, no. Isai, pp. 147–153.

[21] V. Bibhu, K. Roshan, K. B. Singh, and D. K. Singh, "Performance Analysis of Black Hole Attack in Vanet," International Journal of Computer Network and Information Security, vol. 4, no. 11, pp. 47–54, Oct. 2012.

[22] S. Sharma and R. Gupta, "Simulation Study Of Blackhole Attack In The Mobile Ad Hoc Networks," Journal of Engineering Science and Technology, vol. 4, no. 2, pp. 243–250, 2009.

[23] T. Issariyakul and E. Hossain, Introduction to Network Simulator NS2, 2nd ed. Springer, 2012.

[24] Y.-C. Hu, D. B. Johnson, and A. Perrig, "SEAD: secure efficient distance vector routing for mobile wireless ad hoc networks," Ad Hoc Networks, vol. 1, no. 1, pp. 175–192, Jul. 2003.

[25] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks," Wireless Networks, vol. 11, no. 1–2, pp. 21–38, Jan. 2005.

[26] Y. B. Reddy and S. Srivathsan, "Game theory model for selective forward attacks in wireless sensor networks," in Control and Automation, 2009. MED'09. 17th Mediterranean Conference on, 2009, pp. 458–463.

[27] M. Tiwari, K. V. Arya, R. Choudhari, and K. S. Choudhary, "Designing Intrusion Detection to Detect Black Hole and Selective Forwarding Attack in WSN Based on Local Information," in Computer Sciences and Convergence Information Technology, 2009. ICCIT'09. Fourth International Conference on, 2009, pp. 824–828.

[28] T. Sharma, M. Tiwari, P. kumar Sharma, M. Swaroop, and P. Sharma, "An Improved Watchdog Intrusion Detection Systems In Manet," International Journal of Engineering Research & Technology (IJERT), vol. 2, no. 3, pp. 1–4, 2013.

[29] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," in Proceedings of the 6th annual international conference on Mobile computing and networking, 2000, pp. 255–265.

[30] E. Hernandez-Orallo, M. D. Serrat, J. Cano, C. T. Calafate, and P. Manzoni, "Improving Selfish Node Detection in MANETs Using a Collaborative Watchdog," Communications Letters, IEEE, vol. 16, no. 5, pp. 642–645, 2012.

[31] T. Zahariadis, H. C. Leligou, P. Trakadas, and S. Voliotis, "Trust management in wireless sensor networks," European Transactions on Telecommunications, vol. 21, no. 4, pp. 386–395, 2010.

[32] O. Khalid, S. U. Khan, S. A. Madani, K. Hayat, M. I. Khan, N. Min-Allah, J. Kolodziej, L. Wang, S. Zeadally, and D. Chen, "Comparative study of trust and reputation systems for wireless sensor networks," Security and Communication Networks, vol. 6, no. 6, pp. 669–688, 2013.

[33] M. Chakraborty and N. Chaki, "ETSeM: A Energy-Aware, Trust-Based, Selective Multi-path Routing Protocol," in Computer Information Systems and Industrial Management, 2012, pp. 351–360.