# DEFENCE S&T TECHNICAL BULLETIN

Vol. 5 | Num. 2 | Year 2012 | ISSN 1985-6571

## CONTENTS

## AIMS AND SCOPE

The Defence S&T Technical Bulletin is the official technical bulletin of the Science & Technology Research Institute for Defence (STRIDE). It contains articles on research findings in various fields of defence science & technology. The primary purpose of this bulletin is to act as a channel for the publication of defence-based research work undertaken by researchers both within and outside the country.

## WRITING FOR THE DEFENCE S&T TECHNICAL BULLETIN

Contributions to the bulletin should be based on original research in areas related to defence science & technology. All contributions should be in English.

## PUBLICATION

The editors' decision with regard to publication of any item is final. A paper is accepted on the understanding that it is an original piece of work which has not been accepted for publication elsewhere. Contributors will receive two complimentary copies of the issue in which their work appears.

## PRESENTATION OF MANUSCRIPTS

The format of the manuscript is as follows:

a) Page size B5 (JIS)
b) MS Word format
c) Single space
d) Justified
e) In Times New Roman ,11-point font
f) Should not exceed 20 pages, including references
g) Texts in charts and tables should be in 10-point font.

Please e-mail the manuscript to:

1) Dr. Zalini bt Yunus (zalini.yunus@stride.gov.my)
2) Dr. Mahdi bin Che Isa (mahdi.cheisa@stride.gov.my)
3) Dinesh Sathyamoorthy (dinesh.sathyamoorthy@stride.gov.my)

The next edition of the bulletin is expected to be published in April 2013. The due date for submissions is 20 February 2013. **It is strongly iterated that authors are solely responsible for taking the necessary steps to ensure that the submitted manuscripts do not contain confidential or sensitive material.**

The template of the manuscript is as follows:

# TITLE OF MANUSCRIPT

Name(s) of author(s)

Affiliation(s)

E-mail:

## ABSTRACT

*Contents of abstract.*

**Keywords:** *Keyword 1; keyword 2; keyword 3; keyword 4; keyword 5.*

## 1. TOPIC 1

Paragraph 1.

Paragraph 2.

### 1.1 Sub Topic 1

Paragraph 1.

Paragraph 2.

## 2. TOPIC 2

Paragraph 1.

Paragraph 2.

**Figure 1: Title of figure.**

**Table 1: Title of table.**

| Content | Content | Content |
|---|---|---|
| Content | Content | Content |
| Content | Content | Content |
| Content | Content | Content |

Equation 1 (1)
Equation 2 (2)

## REFERENCES

Long lists of notes of bibliographical references are generally not required. The method of citing references in the text is 'name date' style, e.g. 'Hanis (1993) claimed that...', or '…including the lack of interoperability (Bohara *et al.*, 2003)'. End references should be in alphabetical order. The following reference style is to be adhered to:


Books

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

Book Chapters

Goodchild, M.F. & Quattrochi, D.A. (1997). Scale, multiscaling, remote sensing and GIS. *In* Quattrochi, D.A. and Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. Lewis Publishers, Boca Raton, Florida, pp. 1-11.

Journals / Serials

Jang, B.K. & Chin, R.T. (1990). Analysis of thinning algorithms using mathematical morphology. *IEEE T. Pattern Anal.*, **12**: 541-550.

Online Sources

GTOPO30 (1996). *GTOPO30: Global 30 Arc Second Elevation Data Set*. Available online at: http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html (Last access date: 1 June 2009)

Unpublished Materials (e.g. theses, reports and documents)

Wood, J. (1996). *The Geomorphological Characterization of Digital Elevation Models*. PhD Thesis, Department of Geography, University of Leicester, Leicester.

# CHANGES IN BODY MASS INDEX (BMI) AND BODY COMPOSITION IN MALAYSIAN ARMY (MA) PERSONNEL FOLLOWING TWO WEEKS OF STRENUOUS MILITARY TRAINING

Aznida Yusuf[1*], Mohd Ismail Noor[2], Norimah A.Karim[2] & Ridwan Yahaya[1]

[1]Science and Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia
[2]Faculty of Allied Health Sciences, National University of Malaysia (UKM), Malaysia

*Email: aznida.yusuf@stride.gov.my

## ABSTRACT

*The purpose of this study is to evaluate the body composition of Malaysian Army (MA) personnel and to determine changes in body composition following two weeks of strenuous military training. A total of 40 male subjects, who were apparently healthy, passed the Army Fitness Assessment Test (AFAT) and aged between 18 to 35 years were selected. Body mass index (BMI) was determined as weight / height$^2$ while body composition was measured using the bioelectric impedance analysis (BIA) method. The mean BMI of the subjects (23.0 ± 3.0 kg.m$^{-2}$) was found to be within the normal range. The two-week exercise resulted in a decrease of total body fat and increase in lean body mass due to higher physical activity. Body fat content showed a significant reduction ($p < 0.05$) between Days 1 and 14 of the training.*

**Keyword:** *Body composition changes; military training; bioelectric impedance analysis (BIA) method; physical performance; obesity.*

## 1. INTRODUCTION

The Malaysian Army (MA) is building the capacity to have a knowledgeable, smart, mentally prepared and physically strong army to defend the nation. Army personnel must be prepared at all junctures, and participate in combat missions to fulfil all the physical demands of their profession. Studies have identified effective recruitment, better training, and skills with adequate experience as the main aspects which are to be addressed as 'a force to win' (Booth *et al.*, 2002). In addition, response to immediate emergency conditions requires proper health and fitness of army personnel. Army personnel are also required to be physically and psychologically fit with appropriate nutritional and immunological status. It is

pertinent to mention that the relationship of body composition to performance of physical tasks be considered as an important component for ascertaining physical fitness. Body fat in excess of the acceptable amount may impair physical performance as it increases the mass of the individual and hampers the movement of the body (Duangporn, 2012).

Determination of body composition has relied on various methodologies such as hydrostatic weighting, body plethysmography, skinfolds, dual energy X-ray absorptiometry (DEXA), anthropometry, and electrical impedance (Williams, 2002). The skinfold and bioelectrical impedance analysis (BIA) methods are simpler methods which can be used in field settings (Prentice & Jebb, 2001). Body mass index (BMI) has often been chosen as a surrogate measure of body fatness because of its simplicity and the association of BMI with mortality (Poston & Foreyt, 2002). BMI reasonably correlates to body fatness and health risks. However, BMI has its own limitation as it does not incorporate any measure of body composition (Arroyo *et al.*, 2004). As a consequence, many muscular individuals can be misclassified as overweight instead of normal weight (Heinrich *et al.*, 2008).

The MA uses BMI as an indicator to screen prospective applicants prior to entry into military services. For those already in service, their BMI values are monitored every six months while their fitness performances are determined using the Basic Army Fitness Test (BAFT) (Army Training Headquaters, 1999). Failure to comply with BMI standards may result in deferment of promotion.

In order to keep the army personnel active in their physical routines, it is important to know their BMI, body composition and physical fitness (Razalee *et al*., 2010). To this end, we embarked on the present study with the main aim of evaluating the body composition of the army personnel and to determine the changes in body composition following two weeks of strenuous military training.

## 2. METHODOLOGY

### 2.1 Subjects and Study Location

Officers attending the Platoon Tactics Course in the Malaysian Army Combat Training Centre (PULADA) were screened for this study. Inclusion criteria included age between 18 to 35 years, apparently healthy, and passing the Army Fitness Assessment Test (AFAT). A total of 40 subjects fulfilled the inclusion criteria. These subjects were briefed on the purpose, risks, benefits and procedures of the study prior to signing a statement of consent. The study's procedures were approved by the Army Headquarters Operation and Training Branch, and Science and Technology Research Institute for Defence (STRIDE).

The subjects were involved in a two-week military training, which included seven days of conventional warfare training in Kampung Sedili Kecil, Kota Tinggi, Johor, and seven days of military training in Kem PULADA, Ulu Tiram. Johor. Socio-demographic data was collected using a self-administered questionnaire, which included questions such as race, marital status, education status and rank. Throughout the training, meals were given according to the normal armed forces ration (NAFR) of approximately 2700 kcal/day, with energy balance recorded.

## 2.2 Physical Training Regimen

In the first week of the training, the subjects were exposed to real war scenarios. The war was divided into three phases, i.e., defend, withdraw and attack. After the troops arrived and the site was established, they were divided into different platoons. Each platoon established their own post. In the three-day defence phase, the subjects were engaged in digging and patrolling. A slit trench was dug up to 6 m deep around the post. During the next two days, subjects were ordered for immediate withdrawal of 6 to 10 km away from the post. Their main activities consisted of walking, running and organising attacks with a carrying load of 13 kg. In the following two days, the subjects were engaged in the attack phase that simulated a true war scenario.

In PULADA (i.e., the eighth day), the day started with physical activity as early as 6.45 am, followed by marching. The subjects were engaged in theory classes in the morning and practical in the evening. The subjects for the theory class included leadership and management, battle drills, patrolling technique, and principles of defence. The classes ended at 2.30 pm. The practical sessions in the evenings consisted of firing technique, physical fitness and other military related tasks. Dinner was served at 8.00 pm and class continued until 10.00 pm.

## 2.3 Body Composition Measurement

The weight and height of the subjects, while wearing minimal clothing, were measured using a Tanita THD 306 digital weighing scale (Tanita, Japan) and a Seca 208 body meter (Seca, Germany) respectively. For each measurement, weight was recorded to the nearest 0.1 kg and height to the nearest 0.1 cm. BMI in $kg.m^{-2}$ was then calculated. WHO's (2000) guidelines were used to categorise individuals with BMI below 18.5 $kg.m^{-2}$ as underweight, 18.5 to 25 $kg.m^{-2}$ as normal, 25 to 30 $kg.m^{-2}$ as pre-obese, and above 30 $kg.m^{-2}$ as obese.

Body composition was measured using the bioelectric impedance analysis (BIA) method (Bodystat® 1500 from Bodystat Ltd, UK). The instrument uses a single electric frequency of 50 kHz and the Lukaski prediction equation to estimate

percentage body fat from a calculated estimate of total body water (Lukaski *et al.*, 1985). The subjects were to arrive at the measurement area early in the morning following 4-5 hours of fasting. No physical activity was allowed 12 hours before measurement. Alcohol and caffeine intake was restricted during this study. Identical measurement protocols were used for pre- and post-training.

### 2.4 Statistical Analysis

The data was analysed using Statistical Package of Social Sciences (SPSS) 16.0 (SPSS Inc., Chicago, Illinois). Descriptive statistics such as mean, standard deviation and percentage were used to describe variables such as weight, height, body composition and BMI. Independent *t*-test and paired *t*-test were used to compare the difference between groups and among group pre- and post-training. Statistical probability level of $p < 0.05$ was considered significant.

## 3. RESULTS

### 3.1 Socio-Demographic Data

The socio-demographic data of the subjects are shown in Table 1. The subjects were distributed equally by age group (50% in the age group of 18-25 years old and the other 50% in 26-35 years old age group). The majority of the subjects were Malays, bachelors, had completed secondary education and were ranked corporal.

### 3.2 Physical Characteristics and Body Composition at Baseline

The mean age, weight, height and other body composition data of the subjects are shown in Table 2. The mean BMI of the subjects was within the normal range (WHO, 2000).

### 3.3 Body Composition Before, During and After the Military Training

Parameters such as weight, body fat and lean body mass were compared before (Day 0), during (Day 8) and after (Day 14) the training (Table 3). One way analysis of variance (ANOVA) with least significant difference (LSD) post hoc analysis was used to analyse the data. Body fat content showed a significant reduction ($p < 0.05$) between Days 1 and 14, of 2.2 kg (21.46 %). No significant change was observed ($p > 0.05$) between Days 1 and 14 for weight and lean body mass.

**Table 1: Socio-demographic data of the subjects.**

| Factor | *n*=40 | Percentage (%) |
|---|---|---|
| **Age** | | |
| 18-25 | 20 | 50 |
| >25 | 20 | 50 |
| **Race** | | |
| Malay | 32 | 80 |
| Chinese | 2 | 5 |
| Indian | 0 | 0 |
| Others | 6 | 15 |
| **Marital Status** | | |
| Bachelor | 23 | 58 |
| Married | 17 | 42 |
| **Education** | | |
| Degree | 9 | 22 |
| Diploma | 11 | 28 |
| SPM | 20 | 50 |
| **Rank** | | |
| Corporal | 16 | 40 |
| Sergeant | 4 | 10 |
| Second Lieutenant | 9 | 22 |
| Lieutenant | 11 | 28 |

**Table 2: Mean physical characteristics and body composition of the subjects.**

| Physical characteristics and body composition | Mean and SD (*n*=40) |
|---|---|
| Age (years) | 28 ± 5 |
| Weight (kg) | 64.5 ± 10.1 |
| Height (cm) | 167.1 ± 4.8 |
| BMI ($kg.m^{-2}$) | 23.0 ± 3.0 |
| Body fat (kg) | 10.2 ± 3.5 |
| Body fat (%) | 15.8 ± 5.4 |
| Free fat mass (kg) | 54.1 ± 7.4 |
| Free fat mass (%) | 83.7 ± 6.4 |
| Total Body Water (kg) | 39.64 ± 4.07 |

**Table 3: Body composition measurement before, during and after the training.**

| Parameter measured | Weight (kg) | Body fat (kg) | Lean body mass (LBM) (kg) |
|---|---|---|---|
| Day 0 (kg) mean ± SD | 64.5 ± 10.1 | 10.2 ± 3.5 | 54.1 ± 7.4 |
| Day 8 (kg) mean ± SD (Δ8-0) | 64.3 ± 10.1 (-0.23) | 9.7 ± 3.1 (-0.5) | 54.6 ± 8.0 (0.52) |
| Day 14 (kg) mean ± SD (Δ14-0) | 63.3 ± 9.8 (-1.19) | 8.0 ± 2.8 (-2.18) | 55.3 ± 7.9 (1.17) |

## 4. DISCUSSION

The results of weight and height in this study are similar to previous studies conducted for Royal Malaysian Navy (RMN) (Razalee *et al.*, 2010) and MA (Isa *et al.*, 1991; Ismail *et al.*, 1996) personnel. Body fat percentage was found to be normal (16.5-20.5%) according to USDHHS (1996), but results obtained were higher from the studies conducted at Lumut (23.5 ± 6.8 %) and Kuantan (24.2 ± 6.0 %) RMN Bases. The mean BMI data for the subjects in this study, and among RMN and MA personnel in the previous studies was also within the normal range.

While comparing with other countries (Table 4), this study showed smaller height and weight of MA personnel as compared to previous studies conducted for Belgian (Mullie *et al.*, 2008), United States (US) (Graham *et al.*, 2000; Nindl *et al.*, 2007) and Norwegian (Hoyt *et al.*, 2006) military personnel. The reasons for the smaller height and weight of the personnel maybe explained by the geographical distribution of the subjects.

The majority of personnel from the US Army Rangers (Nindl *et al.*, 2007), and US Navy and Marine Corps (Graham *et al.*, 2000) were found to be in the pre-obese range, with mean BMI index of 25.6 ± 4.2 $kg.m^{-2}$ and 25.5 ± 0.41 $kg.m^{-2}$ respectively. However, body fat percentage was 18.5 ± 4.4 % for the US Army Rangers and 19.8 ± 0.7 % for the US Navy and Marine Corps. This percentage falls within normal range and hence, they are not exactly pre-obese. While BMI index is a good guideline for a moderate sized person, one of its weaknesses is that it can misclassify a muscular person to the pre-obese category (Heinrich *et al.*, 2008).

**Table 4: Comparison of baseline data between local and foreign soldiers.**

| Population | *n* | Age (year) | Height (m) | Weight (kg) | BMI ($kg.m^{-2}$) | Body fat (kg) | Body fat (%) | Free fat mass (kg) |
|---|---|---|---|---|---|---|---|---|
| Army – Base Camp (Isa *et al.,*1991) | 35 | 20.5 ± 1.2 | 165.7 ± 5.8 | 56.5 ± 4.5 | 20.4 ± 1.6 | 8.2 ± 3.2 | 14.5 ± 2.9 | 48.0 ± 3.6 |
| Army-Training Camp (Isa *et al.*, 1991) | 35 | 31.8 ± 2.3 | 164.4 ± 17.1 | 63.9 ± 8.1 | 22.2 ± 3.5 | 11.1 ± 2.4 | 17.4 ± 4.4 | 52. 8 ± 4.8 |
| Army-Base Camp (Ismail *et al.*, 1996) | 20 | 25.1 ± 3.8 | 167.3 ± 6.2 | 61.1± 5.8 | 21.9 ± 1.5 | 10.0 ±3.4 | 16.4 ± 3.4 | 51.0 ± 4.0 |
| US Navy and Marine Corps (Graham *et al.*, 2000) | 225 | 25-34 | 178.2 ± 8.2 | 81.3 ± 1.3 | 25.5 ± 0.4 | - | 19.8 ± 0.7 | - |
| Norwegian military personnel (Hoyt *et al.*, 2006) | 10 | 23.3 ± 1.6 | 177.5 ± 3.7 | 77.6 ± 4.5 | 26.0 | 12.2 ± 2.7 | 15.7 ± 3.2 | 64.1± 4.1 |
| US Army Rangers (Nindl *et al.*, 2007) | 50 | 24.6 ± 4.1 | 176.1 ± 7.8 | 78.4 ± 8.7 | 25.6 ± 4.2 | - | 18.5 ± 4.4 | - |
| Belgian military candidates (Mullie *et al.*, 2008) | 448 | 19.0 ± 0.8 | 177.5 ± 10.8 | 70.0 ± 10.8 | 22.2 ± 2.8 | - | 17.2 ± 2.8 | - |
| Lumut RMN Base (Razalee *et al.*, 2010) | 706 | 29.5 ± 6.2 | 167.8 ± 5.3 | 68.8 ± 12.0 | 24.4 ± 3.9 | 16.9 ± 7.5 | 23.5 ± 6.8 | 52.0 ± 5.9 |
| Kuantan RMN Base (Razalee *et al.*, 2010) | 95 | 25.4 ± 4.3 | 167.4 ± 5.7 | 64.0 ± 10.4 | 22.7 ± 3.3 | 14.0 ± 6.1 | 24.2 ± 6.0 | 49.6 ± 5.2 |
| Pengerang RMN Base (Razalee *et al.*, 2010) | 108 | 22.5 ± 5.4 | 165.3 ±5.5 | 59.1 ± 8.1 | 21.5 ± 2.2 | 10.2 ± 4.7 | 16.8 ± 5.3 | 48.5 ± 4.3 |
| Present study | 40 | 28.3 ± 5.0 | 167.1 ± 4.8 | 64.5 ± 10.1 | 23.0 ± 3.0 | 10.2 ± 3.5 | 15.8 ± 5.4 | 54.1 ± 7.4 |

There is a positive correlation between BMI and body fat in this study. The coefficient of correlation between BMI and body fat is $r = 0.61$, $p < 0.05$ but Razalee *et al.* (2010) reported a higher coefficient of correlation among RMN personnel, $r = 0.88$. Friedl *et al.* (2002) found that the correlation between BMI and body fat in the US Army was less clear cut because of greater range and variability of muscle mass. A study on US male recruits showed that the correlation between BMI and body fat were the same with US male population, $r = 0.75$ and female recruits, $r = 0.69$ (Robbins & Fonseca, 2001). The present results was closer to that of the US female recruits, and this supports earlier research findings that BMI alone cannot be helpful in determining body fat content (Poston & Foreyt, 2002)

The decrease in body fat was significant ($p < 0.05$) and higher between Days 0 and 14. There was also an increment, but not significant, in lean body mass between Days 0 and 14. The decrease in body fat and an increase in lean body mass was a result of increased physical activity especially during the conventional warfare exercises. Hence, physical fitness profile and muscle strength also increased. This was supported by Nindl *et al.* (2000) study that was conducted based on 18 male and 36 female subjects. They were engaged in physical activity for 5-7 days a week for 24 weeks. The results showed 10% decrease in body fat and 2.2% increase in lean body mass.

Forbes-Ewan *et al.* (2000) reported that Australian military personnel who were on surveillance jobs with loads of 45-65 kg for 10 days showed significant weight reduction between 0.15 and 1.9 kg. Booth *et al.* (2002) in her study on British Army personnel conducting exercises in Sabah found that the subjects lost weight by 4.3 to 5.4%. Neither study reported on increase in lean body mass.

Wilmore *et al.* (1999) in her study on 557 male and female subjects showed that body fat reduced by 3% over a 20-week period of exercise. Lean body mass increment (1.0 kg) was also observed. It should be noted that Friedl *et al.* (1994) reported that loss of total body weight by more than 10% will have a negative impact on health status and work performance.

The findings of the study have their own limitations. The method used to measure body fat is a good field method but it is subject to errors. In spite of this limitation, this study confirmed that the two-week exercise resulted in a decrease in total body fat and increase in lean body mass due to higher physical activity among the subjects.

The data collected provide a baseline data on BMI and body composition of MA personnel. High body fat percentage is of particular interest because it is related to increase in morbidity (Heinrich *et al.*, 2008) hence will reduce individual physical performance. Further studies with more subjects and different BMI groups could give us more information on the association between BMI, body fat

percentage and physical performance. The data will help ensure individual soldier readiness.

## 5. CONCLUSION

In conclusion, this study showed smaller height and weight of MA personnel as compared to Belgian, US and Norwegian military personnel. The reasons for the smaller height and weight may be explained by the geographical distribution of the subjects, and may have implication on physical training programmes. It was found that the majority of the subjects were in the normal category according to BMI classification. The two-week exercise resulted in a decrease in total body fat and increase in lean body mass due to higher physical activity by the subjects. This study also confirmed the adverse effect of reduced in percent body fat on fitness profile.

## ACKNOWLEDGEMENT

The authors are grateful to the military personnel who participated in the study. The kind help received from the Malaysian Army Combat Training Centre (PULADA) and Science and Technology Research Institute for Defence (STRIDE) is highly acknowledged.

## REFERENCES

Army Training Headquarters (1999). Army Fitness Assessment Test (AFAT) (MK TD/G1/3701/(D&P)). Army Training Headquarters, Malaysian Army, Kuala Lumpur.

Arroyo, M., Rocandio, A.M., Ansotegui, L., Herrera, H., Salces, I. & Rebato, E. (2004). Comparisons of predicted body fat percentage from antropometric methods and from impedance in university students. *Br. J. Nutr.*, **92**: 827-832.

Booth, C.K., Coad, R.A. & Roberts, W. (2002). *The Nutritional, Physiological and Psychological Status of a Group of British Sappers after 23 days of Adventure Training in the Hot Wet Tropics*. DSTO Technical Report DSTO-RR-0243. CBRN Defence Centre. DSTO Platforms Sciences Laboratory, Victoria, Australia.

Chumlea, W.C., Guo, S.S., Kuczmarski, R.J., Flegal, K.M., Johnson, C.L., Heymfield, S.B., Lukaski, H.C., Friedl, K. & Hubbard, V.S. (2002). Body composition estimates from NHANES III bioelectrical impedance data. *Int. J. Obes.*, **26**: 1596-1609.

Department of Defence (DOD) (2000). *Australian Defence Force Health Status Report*. Department of Defence, Canberra.

Duangporn, T.N., Maneerat, C. & Vikit, K. (2012). Body mass index and percentage of body fat determined physical performance in healthy personnel. *Asian Biomed.,* **6**: 313-318

Frankenfield, D.C., Rowe, W.A., Cooney, R.N., Smith, J.S. & Baker, D. (2001). Limits of body mass index to detect obesity and predict body composition. *Nutrition*, **17**: 26-30.

Friedl, K.E. & Leu, J.R. (2002). Body fat standards and individual physical readiness in a randomized army sample: screening weight, methods of fat assessment and linkage to physical fitness. *Milit. Med.*, **167**: 994-1000.

Friedl, K.E., Moore, R.J. & Martinez-Lopez, L.E. (1994). Lower limits of body fat in healthy active men. *J. Appl. Physiol.*, **77**: 933-940.

Forbes-Ewan, C.H., Booth, C.K., Coad, R.A., Thompson, G.F., Niro, P.J., Davies, P.S.W. & Atkin, L.M. (2000). Effect of medium-term feeding with combat ration packs on military performance. *2000 Pre-Olympic Congress*. Australia.

Gantt, C.J., Neely, J.A., Villafana, I.A., Chun, C.S. & Gharabaghli, S.M. (2008). Analysis of weight and associated health consequences of the active duty staff at a major naval medical centre. *Mil. Med.*, **173**: 434-440.

Goris, A.H. & Westerterp, K.R. (2008). Physical activity, fat intake and body fat. *Physiol. Behav.*, **94**: 164-168.

Graham, W.F., Hourani, L.L., Sorenson, D. & Yuan, H. (2000). Demographic difference in body composition of Navy and Marine Corps personnel: Findings from the perception of wellness and readiness assessment. *Mil. Med.*, **165**: 60-69.

Harrison, L., Brennam, M.A. & Levine, A.M. (2000). Physical activity patterns and body mass scores among military service members. *Am. J. Health Promot.*, **15**: 77-80.

Heinrich, K.M., Jitnarin, N., Suminski, R.R., Barkel, L., Hunte, C.M., Alvarez, L., Brundige, A.R., Peterson, A.L., Foreyt, J.P.. Haddock, C.K., Poston, W.S.C. (2008). Obesity classification in military personnel: A comparison of body fat, waist circumference and body mass index measurements. *Mil. Med.*, **173**: 67-73.

Heyward, V. & Wagner, D. (2004). *Applied Body Composition Assessment, 2nd Ed.* Human Kinetics, Illinois, pp. 51-56.

Hales, J.R.S., Hubbard, R.W. & Gaffin, S.L. 1996. Limitation of heat tolerance. Dlm. Fregly, M.J. & Blatteis, C.M. (pynt). *Handbook of Physiology, Section 4: Environmental Physiology,* New York: Oxford University Press, pp. 285-355.

Heinrich, K.M., Jitnarin, N., Suminski, R.R., Berkel, L., Hunter, C.M., Alvarez, L., Brundige, A.R., Peterson, A.L., Foreyt, J.P., Haddock, C.K. & Poston,

W.S.C. (2008). Obesity classification in military personnel: A comparison of body fat, waist circumference and body mass index measurement. *Mil.Med.*, **173**: 67-73

Hoyt, R.W., Opstad, P.K., Hauges, A., Delany, J.P., Cymerman, A. & Friedl, K.E. (2006). Negative energy balance in male and female rangers: effects of 7 d of sustained exercise and food deprivation. *Am. J. Clinc. Nutr.*, **83**: 1068 – 1075.

Isa, M. (1991). *Kajian Keperluan Tenaga Di Kalangan Anggota Angkatan Tentera Malaysia.* MSc. Thesis, Universiti Kebangsaan Malaysia, Bangi, Selangor.

Ismail Mohd Noor, Isa Mansor & Janudin, A. (1996). Energy requirements of Malaysian soldiers in a base camp. *Mal. J. Nutr*., **2**:168-174.

Lindquist, C.H. & Bray, R.M. (2001). Trends in overweight and physical activity among US military personnel, 1995 – 1998. *Prev. Med*., **32**: 57 – 65.

Mullie, P., Vansant, G., Hulens.M., Clarys, P. & Degrave, E. (2008). Evaluation of body fat estimated from body mass index and impedance in Belgian male military candidate: comparing two method for estimating body composition. *Mil. Med.*, **173**: 266-270.

Nelson, H.A., Thomson, C.A. & Klein, N.L. (2005). Effects of healthy lifestyle intervention versus healthy lifestyle intervention plus knowledge of measured resting energy expenditure on weight and body fat in Air Force personnel. *J. Am. Diet. Assoc.*, **105**: 10.

Nindl, B.C., Barnes, B.R., Alemany, J.A., Frykman, P.N., Shippee, R.L. & Friedl, K.E. (2007). Physiological consequences of U.S. Army ranger training. *Med. Sci. Sports Exerc.*, **39(8)** : 1380-1387.

Nindl, B.C., Harman, E.A., Marx, J.O., Gotshalk, L.A.,Frykman, P.N., Lammi, E., Palmer, C. & Kraemer, W.J. (2000). Regional body changes in women after 6 months of periodized physical training. *J. Appl. Physiol.*, **88**: 2251-2259.

Poston W.S.C. & Foreyt J.P. (2002). Body mass index: uses and limitations. *J Strength Cond Res.*, **24**: 15-27.

Prentice, A.M. & Jebb, S.A. (2001). Beyond body mass index. *Obes. Rev.*, **2**: 141-147.

Razalee, S., Poh, B.K. & Ismail, M.N. (2010). Body mass index and body composition among Royal Malaysian Navy (RMN) personnel. *J. Defence Secur.*, **1**: 65-82.

Robbins, A.S. & Fonseca, V. (2001). Body mass index and adiposity in active duty military members. *Military Med.*, **166**: 4-5

U.S. Department of Health and Human Services (USDHHS) (1996). *Physical Activity And Health: A Report Of The Surgeon General*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta.

Wagner, D.R. & Heyward, V.H. (1999). Techniques of body composition assessment: A review of laboratory and field methods. *Res. Q. Exercise Sport*, **70**: 135-49

World Health Organization (WHO) (2000). Obesity: Preventing and Managing The Global Epidemic. Report of a WHO consultation. *WHO Tech. Rep. Ser.*, **894**: 1-253.

Wilmore, J.H., Despres, J., Stanforth, P.R., Mandel, S., Rice, T., Gagnon, J., Leon, A.S., Rao, D.C., Skinner, J.S. & Bouchard, C. (1999). Alteration in body weight and composition consequent to 20 week of endurance training: the HERITAGE Family study. *Am. J. Clinc. Nutr.*, **70** : 346 – 352.

Wilmore, J.H. (1996). Increasing physical activity: alteration in body mass and composition. *Am. J. Clinc. Nutr.*, **63** : 456-460.

Williams, M.H. (2002). *Nutrition for Health, Fitness and Sports, 6th Ed.* McGraw-Hill, New York.

Yamane, G.K. (2007). Obesity in civilian adults: potential impact on eligibility for U.S. military enlistment. *Mil. Med.*, **172**: 1160-1165.

# IMPLEMENTATION OF THE BIOLOGICAL AND TOXIN WEAPONS CONVENTION (BTWC) IN MALAYSIA: CHALLENGES AND THE WAY FORWARD

Hanis Haziqah Md Hambali, Megat Hafizal Megat Ramli, Noorliza Hamdan & Zalini Yunus*

Protection & Biophysical Technology Division (BTPB), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*Email: zalini.yunus@stride.gov.my

## ABSTRACT

*Technological advancements in the field of life sciences and biotechnology bring about unprecedented risks of major disease outbreaks, biological warfare and bioterrorism. The Biological and Toxin Weapons Convention (BTWC), albeit instrumental in preventing such occurrences, is insufficient by itself to sustain pertinent safety and security against such risks, what with the lack of a verification regime. The BTWC implementation in Malaysia is thus essential for biological arms control in the country, by which the development and use of biological weapons shall be prohibited, and any biological agent and toxin shall be handled safely and securely. The purpose of this paper is to identify the challenges and obstacles that must be overcome before the Convention can be successfully implemented. Among the biggest challenges are the lack of verification protocols and domestic legislation in the country, lack of awareness among the life sciences community on issues pertaining to the Convention itself and unfamiliarity with the concept of biosecurity, and confidence-building measures (CBM) submission. In order to ensure the effective implementation of the Convention, a variety of substantial measures must be taken towards undertaking the country's full compliance with the Convention.*

**Keywords:** *Biological weapons; Biological and Toxins Weapons Convention (BTWC); laboratory biosafety and biosecurity measures; confidence building measures (CBM); verification mechanisms protocol.*

## 1. INTRODUCTION

In the last few years back, Malaysia has been affected by a number of disease outbreaks within the country, including the Nipah virus, severe acute respiratory syndrome (SARS) and influenza H1N1 (Lee, 2009). These outbreaks have led to the rapid growth of expertise and researchers conducting various studies

regarding vaccine development for outbreaks. However, this has opened up vast opportunities for the threats of biological weapons. Recent advances in microbiology, particularly in the studies of viruses, the manufacturing and production of vaccines, and the development of industrial microbiology, have now made it possible for the production of sophisticated biological weapon systems. There is always the possibility that any research involving biological agents and toxins could have both positive and negative applications, deliberately or not. Such research is known as "dual use research" in which it could be intentionally used in bioterrorism or the pathogen could be inadvertently be released from the laboratory (Fauci & Collins, 2012).

The threat of biological weapons is not new. The Romans were the first to use biological warfare when they placed rotting carcasses in water supplies to make the water undrinkable for opposing troop. In 1346, the invading Tartar army catapulted the bodies of plague victims into the Crimean Peninsula city of Kaffa and infected its citizens. In 1763, British troops under General Jeffrey Amherst gave Delaware Indians blankets used by victims of smallpox, possibly infecting the susceptible native population. Japan contaminated food and released plague-infected ticks during their conflict with China during World War II. The 2001 anthrax letter attacks in the United States infected 22 people and killed five (FAS, 2011).

Malaysia was among the first state parties to sign up to the Biological and Toxin Weapons Convention (BTWC) when it was opened for signature on 10 April 1972 and entered into force on 26 March 1975. The BTWC is the first multilateral disarmament treaty to prohibit the development, production, acquisition, transfer, retention, stockpiling and use of biological and toxin weapons, and is a key element in the international community's efforts to address the proliferation of weapons of mass destruction. Over the intervening years, increasing numbers of states have joined the Convention. According to the recent BTWC membership list in the official website of the United Nations Office at Geneva (UNOG) (UNOG, 2012a), the BTWC has a total of 165 states parties and 12 signatories.

Malaysia has a number of acts which are related to the control of biological weapons, such as the Biosafety Act 2007, Prevention & Control of Infectious Diseases Act 2006, and Strategic Trade Act 2010, but a specific exact act on biological weapons has yet to be introduced. To this end, the Malaysian government is in the process of drafting the Biological Weapons Bill, which is expected to become a part of the country's legislative framework in ensuring the effective implementation of the BTWC. The bill is aimed at enforcing the safe and secure handling of hazardous biological materials, while ensuring that the requirements do not demotivate or short change life science researchers, advocates and business entities in the country (Mohd Najib, 2011).

In devising ways and methods to effectively implement the protocols of the BTWC in Malaysia, it should be noted that there are several challenges that may obstruct the implementation of the Convention. These issues, if not taken into account while implementing the BTWC in Malaysia, may impede such efforts and could even render them practically useless. To this end, this paper is aimed at reviewing these challenges, and the strategies and measures that can be taken to overcome them.

## 2. CHALLENGES IN IMPLEMENTING THE BTWC

### 2.1 Lack of Verification Protocols

Perhaps the biggest challenge in the implementation of the BTWC in Malaysia lies in the absence of a BTWC verification regime. The failure of the BTWC's states parties to fully establish declaration and verification measures, due to US resistance in 2001, has led to the collapse of such proposals during the BTWC Fifth Review Conference. This has rendered prolonged difficulties in verifying and monitoring state parties' compliance with the Convention (Revill *et al*., 2010).

Without any legally binding agreement that would be essential to strengthen the Convention, state parties are not obligated to demonstrate compliance with the treaty. The lack of verification protocols also makes it difficult to confirm whether or not a state party is taking such preventative measures and actions as described under the BTWC provisions (Kahn, 2011). As a consequence, the absence of verification will lead to lack of transparency, confidence and clarification among state parties (Hunger & Zymorzynska, 2011). Therefore, it is difficult for the Convention itself to effectively ensure among states parties that any production, use or stockpiling of biological weapons is effectively prohibited, and that any harmful biological agents and toxins are handled safely and securely.

### 2.2 National Legislation

This challenge hence brings about the second most prominent challenge, which is the absence of BTWC-related legislation in the country. The lack of a BTWC verification regime, which lays profound reason as to why it will prove arduous for Malaysia to implement the BTWC, is made worse by the lack of national implementation measures. Under Article IV of the BTWC provisions, all state parties are encouraged to adopt national legislation to prohibit and prevent the development, production, stockpiling, acquisition or retention of agents, toxins, weapons, equipment and means of delivery (Pearson, 2005). This obligation is further sustained in the Sixth Review Conference (Woodward & Leal, 2006), in

which state parties are called upon to adopt legislative, administrative, judicial and other measures, including penal legislation to enhance domestic implementation of the Convention.

Nonetheless, the level of national implementation is still poor among many state parties. In 23 March 2012, the Verification Research, Training and Information Centre (VERTIC) completed a survey covering 136 countries to get insight into the countries' national laws and regulations for implementing the BWTC. Based on the survey, less than half of those countries have effective national legislation in place (VERTIC, 2012). Without any BTWC-related legislation, it will be difficult for the authorities to implement the Convention in the absence of the ability and authority required to govern and monitor activities involving biological agents and toxins, as well as to restrict and prevent any growth of biological weapons in the country. Having suitable laws or policies in place will ensure that penalties and punishments are given to perpetrators that violate the Convention (Countryman, 2011).

## 2.3 Confidence-Building Measures

The states parties in the BTWC Second Review Convention in 1986 agreed to introduce confidence-building measures (CBM) for enhanced transparency and information sharing capacity, for the purpose of reducing the occurrences of ambiguities, doubts and suspicions (UNOG, 2012b). CBM implementation in Malaysia may however face certain challenges, the most apparent of which stems from its very objective. Since the introduction of CBM in 1986, Malaysia has thus far only submitted the form three times; which were in 2006, 2010, and 2011 (UNOG, 2012c).

While the need for transparency permits gathering of information and monitoring of activities involving biological agents and toxins, it comes at the cost of secrecy. Organisations, institutes and governmental bodies in Malaysia that are secretive and adhere strongly to confidentiality may find CBM as burdensome and may well be in conflict with their interests. Since the forms are to be submitted to the International Supporting Unit (ISU) of the UN, there is fear, either irrational or reasonable, that the information collected may be exploited by irresponsible parties. Such fear holds true especially in the military, considering that most of the information in their possession are strictly confidential and if disclosed, may give adversaries the advantage to initiate or aggravate a military conflict (Kobletz & Chevrier, 2011). This sense of mistrust appears evident among military groups overseas, who would be more vigilant and suspicious of adversaries who urge both sides to be transparent if peace is to be sought.

There may also be competition among groups that prefer their activities and progress to be kept secret, which serves for multiple reasons, among which are to

reduce the predictability of their behaviour as well as the chance of adversaries developing effective counter-strategies (Chevrier & Hunger, 2000). Another point to consider is that life scientists who have to submit CBM forms may find it unnecessarily troublesome, taking into account its length and intricacy. As a result, while CBM forms are to be submitted annually to the ISU, many may choose to advocate against it and may prefer simpler ways, such as only notifications to local authorities, compared to the seemingly exhaustive forms.

## 2.4 Lack of Understanding and Awareness among the Life Sciences Community

There is also the challenge of public awareness in Malaysia, where the significance and urgency to implement the BTWC may not be fully understood within the life sciences community, let alone by the mass public. Considering the name of the Convention may be misleading, the worst case scenario may be where the public perceives the Convention as not for biological arms control, but instead as a cooperative approach between states parties to develop technologies for such weapons. It is also suggested that one of the reasons as to why countries have not joined the BTWC is due to misunderstandings of the Convention, its provisions, and implications (McLaughlin, 2009).

To gain more insight about the awareness of the life sciences community on issues pertaining to the BTWC, the Science and Technology Research Institute for Defence (STRIDE) conducted a survey covering 194 respondents from government institutions, research and academic institutions, private companies, and non-governmental organisations. The participants consisted of researchers, research assistants, biosafety officers and other relevant professions. The survey revealed that 142 respondents or 73.2 % of the total respondents are not familiar with the BTWC. This staggering figure clearly shows the lack of awareness among the life sciences community. The respondents were further requested to indicate their level of awareness on each BTWC provision, and the results are as shown in Table 1.

The survey shows that more than one third of the respondents are not aware of each provision, let along understand them. The overall results of this survey clearly indicate that the life sciences community has very little awareness and understanding on the listed provisions. Some may even deem the issues of biological weapons as trivial and insignificant, claiming that cases involving biological weapons are merely isolated and rare. It should also be worrisome of how unaware the public or even life scientists are on dual-use concerns and the danger they pose as tools for bioterrorism.

In the same survey conducted by STRIDE, the researchers were also asked about their awareness on dual-use concerns involving biological agents and toxins. The

results are as shown in Table 2. A mere 5.8 % of 52 researchers that participated in the survey are extremely aware of the issues, while 19.2 % are slightly aware, 30.8 % are moderately aware, and 25.0 % and 19.2 % are slightly aware and not aware of the issues respectively. Statistical data at the global level also have evinced that there is little evidence to show that life scientists in the UK, Finland, Germany, Japan, Australia and other countries conceive bioterrorism or bioweapon as a threat, consider that advancements and developments in life sciences could contribute to biological threats, or are aware of the BTWC (Dando, 2011).

**Table 1: Awareness level of the Malaysian life sciences community on the BTWC provisions.**

| | | *No response (%) | Fully (%) | Somewhat (%) | Not Aware (%) |
|---|---|---|---|---|---|
| 1. | Never under any circumstances to acquire or retain biological weapons. | 2.6 | 22.2 | 41.2 | 34.0 |
| 2. | To destroy or divert to peaceful purposes biological weapons and associated resources prior to joining. | 3.1 | 13.9 | 43.3 | 39.7 |
| 3. | Not to transfer, or in any way assist, encourage or induce anyone else to acquire or retain biological weapons. | 2.6 | 20.6 | 42.3 | 34.5 |
| 4. | To take any national measures necessary to implement the provisions of the BTWC domestically. | 2.6 | 11.9 | 42.3 | 43.3 |
| 5. | To consult bilaterally and multilaterally to solve any problems with the implementation of BTWC. | 2.6 | 10.3 | 44.3 | 42.8 |
| 6. | To request the UN Security Council to investigate alleged breaches of the BTWC and to comply with its subsequent decisions | 3.1 | 8.2 | 37.6 | 51.0 |
| 7. | To assist States which have been exposed to a danger as a result of a violation of the BTWC | 3.1 | 7.7 | 40.2 | 49.0 |
| 8. | To do all of the above in a way that encourages the peaceful uses of biological science and technology | 2.6 | 16.0 | 44.3 | 37.1 |

*Indicates the percentage of respondents who did not answer the question.

**Table 2: Awareness level of Malaysian life science researchers on dual-use issues.**

| Awareness level | Percentage (%) |
|---|---|
| Not aware | 19.2 |
| Slightly aware | 25.0 |
| Moderately aware | 30.8 |
| Very aware | 19.2 |
| Extremely aware | 5.8 |

Furthermore, there is also the issue of inadequacy in skills and knowledge to implement certain obligations under the Convention among those who regard the obligations, such as effective laboratory biosecurity, as of little importance. This view is supported by the statistical data obtained from the survey conducted by Mancini & Revill (2008) on 57 universities in Europe, through which it found that only three of these universities offered some form of specific biosecurity module in the life science degree courses, and in all cases this module is optional for students

Therefore, the awareness issue in Malaysia must be dealt with considerably, and if not addressed sufficiently and effectively, may leave various parties questioning the purpose and intention of implementing BTWC. It may be to them too extraneous and unnecessarily burdening, and far more than it is necessary to prevent biological warfare and any use of diseases as weapons in Malaysia.

## 2.5 Terminology of Biosecurity is Still New

The terminology of biosecurity is still new in Malaysia, where the focus has thus far been on biosafety measures. Although the terms “biosafety” and “biosecurity” are often used interchangeably, they refer to different issues. Biosafety technologies and procedures aim to prevent accidental infections of biomedical researchers and releases of dangerous pathogens from research laboratories that could endanger public health or the environment. These objectives can be achieved through “biocontainment,” which involves placing impermeable barriers or filters between the infectious agent and the researcher, and between the laboratory and the outside world. Four levels of increasingly stringent biocontainments referred to as Biosafety Levels (BSL) 1 to 4 - are keyed to the lethality and contagiousness of pathogens, and the availability of protective vaccines or therapeutic drugs (Tucker, 2003a).

Biosecurity, in contrast, denotes policies and procedures designed to protect high-consequence microbial agents and toxins, or critical relevant information, against theft or diversion by those who intend to pursue intentional misuse (Callihan, 2002). According to Sandia National Laboratories, a comprehensive laboratory biosecurity comprises of physical protection, personnel reliability, adequate scientific and commercial program oversight, pathogen accountability, transportation security, and information security (Salerno & Koelm, 2002). All these elements are extremely crucial in order to protect the facilities, and the biological agents and toxins inside the facilities against theft or illicit use by someone who has malicious intents to misuse and produce biological weapons.

Hence, Malaysia needs an effective biosecurity system that requires the integration of relevant technologies and procedures. The guidelines should include, as a minimum, the registration and licensing of facilities that work with dangerous biological agents, based on an agreed list of pathogens and toxins that can be readily updated; physical security and access controls at laboratories and culture collections that possess such agents; systems for the control and accounting of listed pathogens and toxins, both in storage and during experiments; background checks on laboratory personnel; and an emergency plan for responding to breaches in security (Tucker, 2003b).

## 3. THE WAY FORWARD

### 3.1 National Implementation of BTWC

As the BTWC remains without a proper verification mechanism, it is substantially crucial that Malaysia undertakes necessary national implementation measures, whereby relevant legislation should be enacted to enforce the Convention (Isla, 2008). The national implementation will serve for multiple purposes, among which is to prohibit the development, use, stockpiling or production of any biological weapons, as well as to ensure that biological agents and toxins are handled safely and securely. The implementation measures that could be adopted by Malaysia are enforcement and inspection, biosafety and biosecurity measures, penal measures to criminalise treaty-prohibited activities, and transfer control (UNOG, 2009). It is also worth noting that while some may claim that establishing policies relevant to the Convention will suffice in upholding its protocols, legislation will prove far more useful as it can be more comprehensive and exhaustive while being far more rigid and stringent, through which appropriate punishment can be given without compromise to transgressors. Via legislation, moreover, authoritative bodies can be established to employ effective implementation of the Convention, and be given the necessary and sufficient functions and power, to conduct inspections for instance, in order to ensure that both the state's and individual's total compliance with the enacted law.

Other than the enacted legislation, submission of CBM forms could also be made compulsory to complement the national implementation of the BTWC. Under the enacted legislation, stakeholders would be required to submit CBM forms as an annual declaration to demonstrate compliance with the legislation and Convention. This strategy could prove to be beneficial to Malaysia as it encourages openness and transparency, builds up confidence and improves trust among countries, and helps the country to establish and enhance relationships with other countries, which will then bring opportunities for collaboration (Kahn, 2011).

To better implement CBM in Malaysia, a National Contact Point can be established to coordinate the compilation and submission of the forms. The National Contact Point shall assist organisations, institutes and governmental bodies in completing the forms through which they will be guided in selecting, compiling and submitting the necessary data and information (UNOG, 2009). The established body could also serve as a screening point to review the CBM forms before forwarding them to the ISU. The body could also be tasked to devise and suggest improvements to the forms so as to reduce any unnecessary burden placed on stakeholders. Aside from that, the body could also work to improve the submission system, whereby an interactive and user-friendly electronic submission system can be developed and utilised in compiling and submitting the forms.

### 3.2 Awareness-Raising and Education

During the BTWC Sixth Review Conference, the state parties agreed that each country must "*promote the development and educational programmes, to promote the development of training and education programs for those granted access to biological agents and toxins relevant to the Convention and for those with the knowledge or capacity to modify such agents and toxins, in order to raise awareness of the risks, as well as of the obligations of States Parties under the Conventions*" (Whitby *et al*., 2011).

Unfortunately, as evidenced by the survey conducted by STRIDE, there is very limited awareness of the Convention amongst Malaysian life scientists. Therefore, Malaysia is obligated to establish certain measures and programmes to educate and raise awareness within the life sciences community on areas such as dual-use concerns, laboratory biosecurity and biosafety, handling of biological agents and toxins, and implementation of the Convention

The programmes should encompass regularly held workshops, surveys, conferences and trainings to provide comprehensive and effective understanding on relevant topics related to the Convention, and to encourage discussions among

the participants (Weller *et al.,* 2010). The target of these programmes shall be the life sciences community in not only academic institutions and associations, but also in private and commercial institutions. They should be made aware of the possible adverse social, environmental, health and security consequences of their research, and that they have both ethical and legal responsibilities in this regard.

Other than implementing such programmes, educational institutes and instructors in Malaysia should consider introducing modules and integrated learning on laboratory biosafety and biosecurity, dual use issues, and the Convention in their teaching and course materials (Whitby & Dando, 2010). They should also take advantage of the internet as a powerful tool by making such materials available online to cater for greater audience, invite learning at all levels of society, and optimise the learning process.

### 3.3 Implementation of Laboratory Biosafety and Biosecurity

Aside from organising certain programmes to create awareness, there is a need for the establishment of a National Focal Point to assist on matters related to the implementation of laboratory biosecurity and biosafety (Fidler & Gostin, 2007). The existence of such a focal point can serve as the role of a coordinator, which coordinates the development and implementation of laboratory biosafety and biosecurity in Malaysia, and conducts necessary surveys and research to further enhance the mechanisms. The role of this body shall include assisting institutions, private companies and governmental bodies in establishing such laboratory biosafety and biosecurity measures and management system in their respective facilities.

The established body shall also assist stakeholders to understand and adopt such measures in order to minimise any risks of infections, or release of biological agents and toxins through proper handling of the materials, safe and secure transfer of materials, records and data maintenance, and use of appropriate equipment and containment levels in handling such agents. The National Focal Point could also provide guidelines and support to stakeholders so as to ensure that the implemented laboratory biosafety and biosecurity management is in compliance with the proper guidelines and international standards. It could also be responsible for monitoring and ensuring compliance with the laboratory biosafety and biosecurity management guidelines (Biosecurity Working Group, 2007).

Additionally, the body could offer assistance in designing and developing high biocontainment facilities, namely the BSL-3 and BSL-4 facilities in Malaysia. It could serve the purpose of assisting stakeholders in ensuring that biocontainment facilities are built and maintained according to the international standards, and

standard operating procedures are developed and followed so as to ensure the safe handling of the biological agents and toxins.

### 3.4 Building Human Capacity

In order to achieve greater success in implementing the BTWC in Malaysia through legislation, education, and implementation of laboratory biosafety and biosecurity, we first need to develop human expertise and capabilities in scientific, technical, and policy matters. Malaysia should continue to engage and develop close collaborations with other countries, and regional as well as international organisations so as to foster exchange of information, knowledge and skills. There are many organisations worldwide that offer training courses and knowledge-sharing programmes related to biosafety and biosecurity issues that are crucial for the implementation of the BTWC in Malaysia. The American Biological Safety Association (ABSA) and Asia-Pacific Biosafety Association (A-PBA) are two well-known associations that provide such training, conferences and seminars. Both organisations also offer training courses on biosafety and biosecurity (Burns *et al*., 2011). ABSA, for example, provides accredited courses where the participants will be certified as a "Registered Biosafety Professional" or "Certified Biosafety Professional".

Therefore, the life sciences community in Malaysia must seize the opportunity to get involved and actively participate in any association or collaboration. In order to accomplish the proposed programmes for the BTWC implementation, such as awareness-raising, and implementing biosafety and biosecurity, we need more experts that can carry out such tasks, provide training and offer assistance. Such collaborations could also be an excellent platform to establish and coordinate training programmes locally and internationally according to our needs, which will equip and train our local expertise with the necessary knowledge and skills to perform their duties effectively. The skills and competencies in such area are fundamental and very much needed in envisaging the success of the implementation.

### 3.5 Resources

Developing such systems and mechanisms in Malaysia will require enormous commitment and significant financial resources. For this reason, it is vital that the government, specifically, serves the role as the pillar in implementing the Convention by giving their utmost support, providing necessary resources, as well as assisting private and non-private bodies and organisations to carry out the obligations of the Convention.

It is proposed that the Malaysian government could support the implementation of the Convention by providing special allocations, funds or grants to be utilised by the institutions, private companies and governmental bodies to improve facilities and maintain laboratories according to the international standards. The government could also assist in establishing laboratory biosafety and biosecurity management in the facilities and conducting training programmes for personnel that are responsible for implementing and monitoring the system in their respective facilities. It is also important that the government provide training and adequate resources (human, financial, skills) in order to carry out necessary tasks related to the enacted BTWC laws and regulations in the country.

### 3.6. Improving Disease Surveillance and Reporting

Under Articles VI, VII and X of the BTWC, all States Parties are encouraged to improve disease surveillance and reporting to enhance the effectiveness of the Convention (Revill *et al*., 2010). Such improvements could deter and minimise any attempt of biological attack, and could ensure prompt response in handling any outbreaks caused by biological agents or toxins. Disease surveillance data could also be used to provide any information or indication of any suspicious use or release of biological agents or toxins.

To ensure the effectiveness of disease surveillance and reporting, the government should encourage research and development of effective surveillance and control of relevant diseases that infect humans, animals and plants, as well as to improve existing reporting mechanisms. Malaysia should also take initiatives to expand its network and cooperate with other countries to promote the exchange of scientific, technical, legal and other information in relevance to the Convention. Such cooperation, regionally or internationally, would benefit Malaysia in improving its disease reporting and surveillance system, in identifying the best mechanisms or standards for handling biological agents and toxins, and in creating better understanding and confidence among the countries.

## 4. CONCLUSION

Despite the absence major threats of bioterrorism in Malaysia, it is completely necessary that the country employs significant measures to prevent such crises from occurring, and to establish ways to mitigate them should they ever occur, considering the number of disease outbreaks which have already taken place. Therefore, the implementation of the BTWC in Malaysia will prove instrumental in preventing biological warfare and any form of bioterrorism in the country, regardless whether it is intentional or unintentional. However, Malaysia will face many challenges in its effort to fully undertake the obligations under the Convention. The first step that should be taken in order to address these

challenges is increasing public awareness. As most of the life sciences community are not aware and unfamiliar with the BTWC, it is very crucial to create awareness programmes to educate and raise their awareness on issues concerning laboratory biosafety and biosecurity, dual-use concerns, and the Convention itself. With awareness comes understanding, and with understanding comes acceptance. This is very critical to ensure that further legislation and domestic laws that will be introduced later is fully accepted by the life sciences community.

The advancement in research and biotechnology has created a sense of urgency for the implementation of the BTWC in Malaysia. Biotechnology is the ultimate double-edged sword. As is the case with most powerful technologies, they can be employed for good or evil. At present, all military and civilian populations throughout the world are vulnerable to a BTWC attack. We remain grossly ill-prepared to respond to an epidemic caused by a novel genetically engineered biological agent. The 20th century was dominated by physics, but recent breakthroughs indicate that the next 100 years likely will be "the Biological Century." In the words of Sir William Stewart as quoted by Reaney (2001):

> "*the First World War was chemical; the Second World War was nuclear; and that the Third World War – God forbid – will be biological.*"

## REFERENCES


Burns, G., Byers, K., Chua, T.M., Sheeley, H., & Goble, B. (2011). Biosafety professionals as stakeholders in the BTWC. *Disarmament Forum: Beyond the BTWC RevCon,* pp. 27-38.

Callihan, D.R. (2002). Regulation and standards. *Appl. Biosaf.,* **7**: 235-249.

Chevrier, M.I. & Hunger, I. (2000). Confidence-building measures for the BTWC: performance and potential. *The Nonproliferation Review/Fall-Winter 2000*, **7**: 24.

Countryman, T. (2011). *Biological Weapons Convention: The Next Five Years.* Available online at: http://www.state.gov/t/isn/rls/rm/175121.htm (Last access date: 5 October 2012).

Dando, M. (2011). *Research Report for the Wellcome Trust Project on 'Building a Sustainable Capacity in Dual-use Bioethic'.* University of Bradford, Bradford, UK.

Fauci, A.S. & Collins, F.S. (2012). Benefits and risks of influenza research: Lessons learned. *Sci.,* **336:** 1522-1523.

Federations of American Scientist (FAS) (2011). *Case Studies in Dual Use Biological Research*. Available online at:

http://www.fas.org/biosecurity/education/dualuse/FAS_Introduction/1_A.html. (Last access date: 15 March 2012).

Fidler, D. & Gostin, L. (2007). *Biosecurity in the Global Age: Biological Weapons, Public Health, and the Rule of Law*. Stanford University Press, UK.

Hunger, I. & Zmorzynska, A. (2011). Verifying and Demonstrating Compliance with the BTWC. *Non-Proliferation Paper No. 5.* pp. 1-12.

Isla, N. (2008). Challenges to the BTWC and some reason for optimism. *International Network of Engineers and Scientists Against Proliferation (INESAP) Information Bulletin No. 28*. pp. 70-74.

Kahn, L.H. (2011). *The Biological Weapon Convention: Proceeding Without a Verification Protocol*. Available online at: http://www.thebulletin.org/web-edition/columnists/laura-h-kahn/the-biological-weapons-convention-proceeding-without-verificatio (Last access date: 10 October 2012.

Koblentz, G.D. & Chevrier, I.S. (2011). Modernizing confidence-building measures for the Biological Weapons Convention. *Biosecure. Bioterror,* **9:** 232-238.

Biosecurity Working Group. (2007). *A Code of Conduct for Biosecurity*. Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands.

Lee, K.C (2009). *Overview of Emerging & Re-emerging Infectious Diseases: A Clear & Present Danger*. Hospital Sungai Buloh, Sungai Buloh, Selangor.

Mancini, G & Revill, J. (2008). *Fostering the Biosecurity Norm: Biosecurity Education for the Next Generation Life Scientists*. The Landau Network-Centro Volta (LNCV), Como, Italy.

McLaughlin, K. (2009). *Building a Global Ban: Why States Have Not Joined the BTWC*. Bioweapon Prevention Project (BWPP), Geneva.

Mohd Najib, A.R. (2011). *Opening Speech of the International Congress on Biosecurity, Biosafely and Biodefence 2011 (BioSSD 2011)*, 19-20 July 2011, Putra World Trade Centre (PWTC), Kuala Lumpur.

Pearson, G.S. (2005). *The Central Importance of Legally Binding Measures for the Strengthening of the Biological and Toxin Weapons Convention (BTWC)*. Weapon of Mass Destruction Commission, Stockholm, Sweden.

Reaney, P. (2001). *Animal Diseases is Reminder of Bioterrorism Danger*. Available online at: http://news.google.com/newspapers?nid=1309&dat=20010904&id=26ktAAAAIBAJ&sjid=jXgFAAAAIBAJ&pg=2646,1794313 (Last access date: 10 October 2012)

Revill. J., Ilchman, K., Mcleish, C., & Nightingale, P. (2010). *Disease Reporting & the BWC*. Harvard Sussex Program, UK.

Salerno, R.M. & Koelm. J.G. (2002). *Biological Laboratory and Transportation Security and The Biological Weapons Convention*. Sandia National Laboratories, Albuquerque, New Mexico.

Tucker, J.B., (2003a). Preventing the misuse of pathogens: the need for global biosecurity standards. *Arms Control Today*, **33:** 3-10.

Tucker, J.B., (2003b). *Biosecurity: Limiting Terrorist Access to Deadly Pathogens*. United States Institute of Peace (USIP), Washington D.C.

UNOG (The United Nations Office at Geneva) (2009). *Guide to participating in the Confidence-Building Measures of the Biological Weapons Conventions*. Available online at: http://www.unog.ch/80256EDD006B8954/(httpAssets)/F925FAB12B61A83DC12576860035E2D6/$file/CBM+guide+pre-production+final+2+Dec.pdf (Last access date: 10 October 2012)

UNOG (The United Nations Office at Geneva) (2012a). *Membership of the Biological Weapons Convention.* Available online at: http://www.unog.ch/bwc/cbms (Last access date: 27 February 2012).

UNOG (The United Nations Office at Geneva) (2012b). *Confidence-Building Measures Returns*. Available online at: http://www.unog.ch/__80256ee600585943.nsf/(httpPages)/4fa4da37a55c7966c12575780055d9e8?OpenDocument&ExpandSection=24#_Section24. (Last access date: 25 February 2012).

UNOG(The United Nations Office at Geneva) (2012c).*Confidence-Building Measures.* Available online at http://www.unog.ch/bwc/cbms (Last access date: 27 February 2012).

VERTIC (Verification Research, Training and Information Centre). (2012) . *National Implementation of the BWC: Statistics Based on VERTIC's Legislation Survey (as at 23 March 2012)*. Verification Research, Training and Information Centre (VERTIC), London.

Weller, R.E., Burbank, R.L, & Mahy, H.A. (2010). *Outreach and Education in the Life Sciences: A Case Study of the U.S. Department of Energy National Laboratories*. Pacific Northwest National Laboratory, Richland, Washington D.C.

Whitby, S. & Dando, M. (2010). Biosecurity awarenessraising and education for life scientists: What Should be Done Now. *In* Rappert, B. (Ed.), *Education and Ethics in the Life Sciences: Strengthening the Prohibition of Biological Weapons.* Australian National University Press, Canberra, pp. 179-197.

Whitby, S., Bollaert, C., & Dando, M. (2011) Article IV: National Implementation: Education, outreach, and codes of conduct. *In* Pearson, G.S, Sims, N.A., & Dando, M. (Eds.), *Strengthening the Biological Weapon Convention: Key Point for the Seventh Review Conference.* University of Bradford, Bradford, pp. 113- 140.

Woodward, A & Leal. R.E. (2011). National implementation measures for the Biological Weapons Convention. *Regional Workshop on National Implementation of the BTWC*, 27- 28 June 2011, Manila.

# EVALUATION OF GLOBAL POSITIONING SYSTEM (GPS) PERFORMANCE DURING SIMPLISTIC GPS SPOOFING ATTACKS

Dinesh Sathyamoorthy*, Mohd Faudzi Muhammad, Nor Irza Shakhira Bakthir, Siti Robiah Abdul, Shalini Shafii, Aliah Ismail, Lim Bak Tiang, Zainal Fitry M Amin, Mohd Rizal Ahmad Kamal, Siti Zainun Ali & Mohd Hasrol Hisam M Yusoff

Instrumentation and Electronics Technology Division (BTIE), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*E-mail: dinesh.sathyamoorthy@stride.gov.my

## ABSTRACT

*This study is aimed at evaluating Global Positioning System (GPS) performance during simplistic GPS spoofing attacks, whereby spoofing is conducted using a standalone GPS simulator, which at present poses the greatest near-term threat. The study is conducted on two handheld GPS receivers; Garmin GPSmap 60CSx (evaluated GPS receiver) and Garmin GPSmap 60CS (reference GPS receiver). Both receivers employ the GPS L1 coarse acquisition (C/A) signal. Varying minimum spoofing signal power levels, times between position fix lost and spoofing, and probable error patterns are observed for different dates and times. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over time, resulting in GPS performance being time dependent. Variation in other GPS error parameters, including ionospheric and tropospheric delays, satellite clock, ephemeris and multipath errors, and unintentional signal interferences and obstructions, could have also resulted in the variation of GPS performance. As the spoofing signal power level is increased, probable error values increase due to decreasing carrier-to-noise density ($C/N_0$) levels for GPS satellites tracked by the receiver. For all the readings, the highest probable errors occur at the minimum spoofing power levels. After spoofing takes place, the probable errors reduce to levels that are lower as compared to prior to transmission of the spoofing signal, as at this point, the spoofing signal power levels are relatively large, resulting in high $C/N_0$ levels and hence, improved accuracy.*

**Keywords:** *Global Positioning System (GPS) spoofing; position dilution of precision (PDOP); probable errors; location fix loss; signal synchronisation.*

## 1. INTRODUCTION

There is a steady growth in the entrenchment of Global Navigation Satellite Systems (GNSS) in current and upcoming markets, having penetrated various consumer products, such as cell phones, personal navigation devices (PNDs), cameras and assimilation with radio-frequency identification (RFID) tags, for various applications, including navigation, surveying, timing reference and location based services (LBS). While the Global Positioning System (GPS), operated by the US Air Force (USAF), is the primarily used GNSS system worldwide, the upcoming Galileo and Compass systems, and the imminent conversion of *Global'naya Navigatsionnaya Sputnikovaya Sistema* (GLONASS) signals from frequency division multiple access (FDMA) to code division multiple access (CDMA) look set to make multi-satellite GNSS configurations the positioning, navigation & timing (PNT) standard for the future.

However, many GNSS users are still not fully aware of the vulnerabilities of GNSS systems to various error parameters, such as ionospheric and tropospheric delays, satellite clock, ephemeris and multipath errors, satellite positioning and geometry, and signal interferences and obstructions. These error parameters can severely affect the accuracy of GNSS readings, and in a number of cases, disrupt GNSS signals (Volpe 2001; Kaplan & Hegarty 2006; Dinesh 2009, 2011; Last 2010; Schwartz 2010; RAE 2011; Schue 2012).

Spoofing refers to forging and transmission of false navigation messages in order to manipulate the navigation solutions of GNSS receivers. Spoofing signals can be generated by GNSS simulators, equipment which is available today. The received power of the spoofing signal should exceed that of the legitimate signal, this being essentially a form of jamming. The receiver then operates with the forged signal as the input and computes the location induced by the spoofer. Spoofing is more sinister than intentional jamming because the targeted receiver cannot detect a spoofing attack and hence, cannot warn users that its navigation solution is untrustworthy. While spoofing is more difficult to achieve than jamming, in many cases even if a spoofer is not fully successful, it can still create significant errors and jam GNSS signals over large areas (Volpe, 2001; Johnston & Warner, 2004; Papadimitratos & Jovanovic, 2008; Humphreys *et al.*, 2009; Last, 2010; Jones, 2011; Wesson *et al.*, 2012; Scott, 2012).

A number of GNSS simulators have been designed for legal purposes such as user training, system maintenance, vehicle motion simulation and radio frequency interference (RFI) operability tests, such as conducted in Dinesh *et al.* (2012a, b). However, in the wrong hands, these GNSS simulators can be used to conduct illegal spoofing. Furthermore, GNSS simulators can be built with relatively low cost equipment, as demonstrated by Rogers (1991), Johnston & Warner (2004), Humphreys *et al.* (2008), Hanlon *et al.* (2009) and Nicola *et al.* (2010). Hanlon *et al.* (2009) and Montgomery *et al.* (2009) classify GNSS

spoofers into three categories, simplistic, intermediate and sophisticated, depending on their complexity and level of robustness required to the associated counter-spoofing measures.

While GNSS spoofing has not yet emerged as a major threat, it represents a growing risk, with a number of successful spoofing experiments on GNSS receivers being conducted, whereby the respective receivers failed to detect the presence of such attacks (Warner & Johnston, 2002; Humphreys *et al.*, 2008; Motella *et al.*, 2010; Cavaleri *et al.*, 2010; Tippenhauer *et al.*, 2011; Shepard & Humphreys, 2011; Wesson *et al.*, 2012). In addition, studies have been conducted to demonstrate the vulnerabilities of GNSS-based systems for unmanned aerial vehicle (UAV) navigation and power grid time synchronisation to spoofing (Shepard *et al.*, 2012). Hence, the development of various GNSS counter-spoofing technologies has received significant attention (Key, 1995; Wen *et al.*, 2005; Papadimitratos & Jovanovic, 2008; Humphreys *et al.*, 2009; Montgomery *et al.*, 2009; Ledvina *et al.*, 2009; Nielsen *et al.*, 2010; Pozzobon *et al.*, 2011; Jones, 2011; Daneshmand *et al.*, 2011; Wesson *et al.*, 2012; Sherman *et al.*, 2012; Scott, 2012).

This study is aimed at evaluating GPS performance during simplistic GPS spoofing attacks, whereby spoofing is conducted using a standalone GPS simulator, which at present poses the greatest near-term threat. In this type of spoofing attack, the spoofing signal is not synchronised (in terms of power level, phase, Doppler shift and data content) with the genuine signals received by the target GPS receiver. This could cause the target GPS receiver to temporarily lose position fix lock first, before being taken over by the spoofing signal. Even if the unsynchronised attack could avoid causing loss of lock, it could still cause an abrupt change in the target GPS receiver's time estimate. Rudimentary counter-spoofing measures, such as amplitude and time-of-arrival discrimination, and loss of lock notification, could be used to detect simplistic spoofing attacks. However, many of present civilian GNSS receivers are not equipped with such measures, and hence, are vulnerable to simplistic spoofing attacks (Humphreys *et al.*, 2008; Hanlon *et al.*, 2009; Montgomery *et al.*, 2009; Jones, 2011; Wesson *et al.*, 2012).

The spoofing signals in this study are generated using an Aeroflex GPSG-1000 GPS simulator (Aeroflex, 2010), which was procured by the Science & Technology Research Institute for Defence (STRIDE) for the Tenth Malaysian Plan (RMK10) project entitled *Evaluation of the Effect of Radio Frequency Interference (RFI) on Global Positioning System (GPS) Signals via GPS Simulation*. The study is conducted on two handheld GPS receivers; Garmin GPSmap 60CSx (Garmin, 2007) (evaluated GPS receiver) and Garmin GPSmap 60CS (Garmin, 2004) (reference GPS receiver). Both receivers employ the GPS L1 coarse acquisition (C/A) signal.

## 2. METHODOLOGY

The study was conducted via tests held at the STRIDE Block B car park (Figure 1) in March - May 2012. Trimble Planning's (Trimble, 2012) estimate of GPS satellite coverage indicated that the periods of the tests (UTC 0100 – 0400) coincided with periods of good GPS coverage, with high satellite visibility (generally between 9 to 12 satellites) and low position dilution of precision (PDOP) (generally between 1.2 to 2.5) values. Nevertheless, Trimble Planning only takes into account estimated satellite positions and geometry, and does not consider other sources of GNSS errors. Furthermore, the parameters of elevation cutoff and obstacles were estimated from 30 m resolution terrain models, which do not take into consideration man-made structures, and thereby, are subject to errors.

**Figure 1: Test area located at N 2º 58.056' E 101º 48.586' 70m. (Source: Screen capture from Google Earth)**

It should also be noted that the PDOP values provided by Trimble Planning are best case estimates of GNSS coverage in a particular region. The actual PDOP values obtained by a particular GNSS receiver are dependent on its sensitivity and the strength of GNSS signals received.

The test was conducted using the setup shown in Figure 2. The spoofing signal generated by the GPS simulator was transmitted via a GPS Source A11XLV GPS amplifier (GPS Source, 2006) and a GPS Source L1P GPS passive antenna (GPS Source, 2007). GPS accuracy, in terms of horizontal probable error (HPE), vertical probable error (VPE) and estimate probable error (EPE), was measured using GPS Diagnostics v1.05 (CNET, 2004). An Advantest U3751 spectrum analyser (Advantest, 2009) was employed to monitor for possible unintentional interference signals.

The almanac data for the period of the test was imported into the GPS simulator via its internal GPS receiver. The spoofing signal was set for position of N 2º 58' E 101º 48' 80m (approximately 1 km from the test area), while the time was set at the simulator's GPS receiver's time.

Once a position fix was obtained with the GPS receiver, transmission of the spoofing signal was started at power level of -150 dBm and increased by increments of 3 dBm. The power level at which loss of position fix occurs and the time required for spoofing to take place were noted.

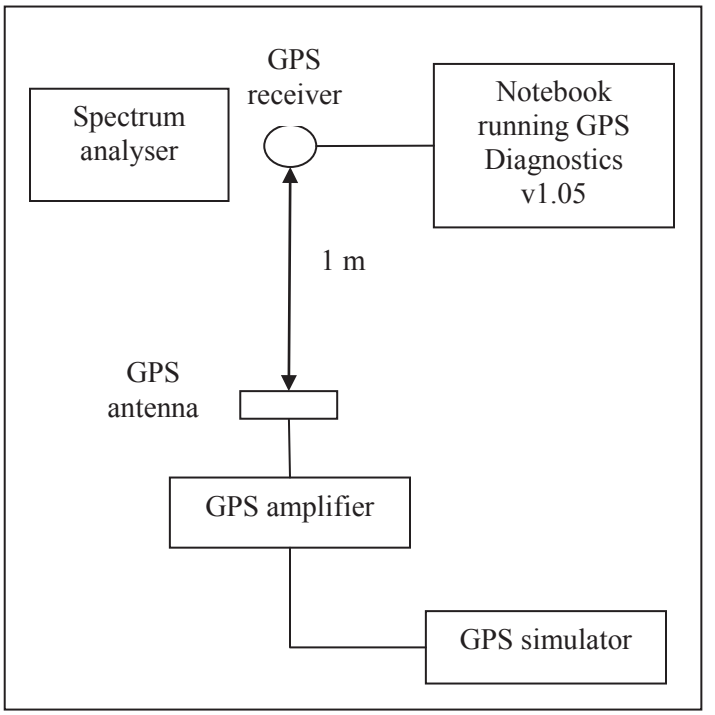



**Figure 2: The test setup employed.**

## 3. RESULTS & DISCUSSION

The results of the tests are shown in Tables 1 and 2. For both the evaluated and reference GPS receivers, varying minimum spoofing signal power levels and times between position fix lost and spoofing are observed for different dates and times. No clear correlation is observed between these two parameters and the corresponding PDOP values. The buildings and trees in the vicinity of the test area could have resulted in the actual PDOP values being significantly different from the estimated values. Furthermore, other GNSS error parameters, including

ionospheric and tropospheric delays, satellite clock, ephemeris and multipath errors, and unintentional signal interferences and obstructions, could have affected GPS coverage during the periods of the tests.

The minimum spoofing signal power level required to cause position fix loss and, subsequently, spoofing is dependent on the received GPS signal power during the tests. During periods of poor coverage, when the received GPS signal power levels are lower, the required minimum spoofing signal powers level would be lower, and vice-versa. It is also observed that the minimum spoofing power levels are significantly higher as compared to the received GPS signal power level (approximately -160 - -130 dBm), as the GPS signal's noise-like C/A code structure, which modulates the L1 signal over a 2 MHz bandwidth, allows for it to be received at low levels of interferences. However, the required minimum spoofing power levels required to cause position fix loss was lower as compared to required minimum interference signal power levels observed during GPS jamming tests conducted both via field evaluations (Dinesh *et al.*, 2009; 2010a,b) and GPS simulation (Dinesh *et al.*, 2012a, b). This occurred as the difference in synchronisation between the genuine and spoofing GPS signals forced the target GPS receiver to recompute its position fix at relatively lower spoofing signal power levels.

**Table 1: The effect of spoofing attacks on the evaluated GPS receiver.**

| Reading | Date (2012) | UTC Time | PDOP | Minimum spoofing signal power level (dBm) | Time between position fix loss and spoofing (s) |
|---|---|---|---|---|---|
| 1 | 29 March | 0104 | 1.72 | -108 | 11 |
| 2 | 29 March | 0240 | 1.72 | -123 | 3 |
| 3 | 3 April | 0105 | 1.62 | -114 | 91 |
| 4 | 3 April | 0234 | 1.72 | -120 | 59 |
| 5 | 2 May | 0108 | 1.47 | -111 | 156 |
| 6 | 2 May | 0244 | 1.27 | -111 | 52 |

**Table 2: The effect of spoofing attacks on the reference GPS receiver.**

| Reading | Date (2012) | UTC Time | PDOP | Minimum spoofing signal power level (dBm) | Time between position fix loss and spoofing (s) |
|---|---|---|---|---|---|
| 1 | 4 April | 0105 | 1.60 | -117 | 347 |
| 2 | 4 April | 0231 | 1.72 | -114 | 266 |
| 3 | 5 April | 0259 | 2.45 | -108 | 276 |
| 4 | 2 May | 0208 | 1.24 | -117 | 379 |
| 5 | 3 May | 0112 | 1.64 | -120 | 283 |
| 6 | 3 May | 0238 | 1.21 | -114 | 412 |

At the minimum spoofing power level, the time between position fix loss and spoofing is dependent on the level of synchronisation, in terms of power level, phase, Doppler shift and data content, between the genuine and spoofing GPS signals. When the both signals are closely synchronised, spoofing occurs very quickly. However, when the signals are largely unsynchronised, position fix loss occurs for a longer period of time as the target GPS receiver has to recompute is position fix.

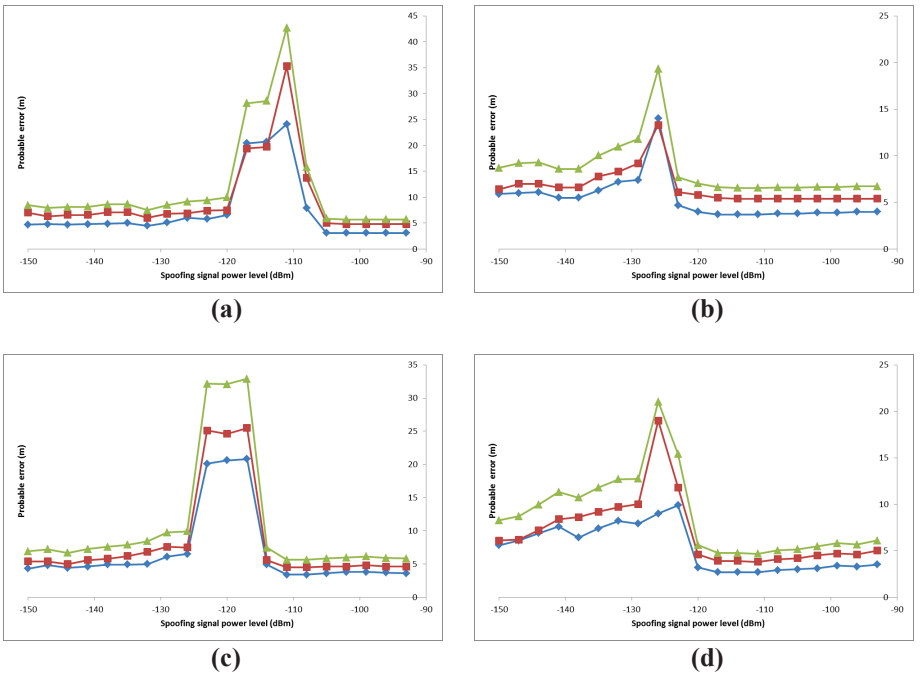
For all the readings, spoofing occurred much faster for the evaluated GPS as compared to the reference GPS receiver. This occurs as the evaluated GPS receiver has significantly better performance characteristics, particularly receiver sensitivity and reacquisition time, that allows it to recompute its position fix much faster as compared to the reference GPS receiver. In other words, the improvements in design that have been made to the evaluated GPS receiver to enhance its performance have, ironically, made it more vulnerable to spoofing.

The probable errors of the GPS receivers during the spoofing attacks are shown in Figures 3 and 4. It is observed that as the spoofing signal power level is increased, probable error values increase due to decreasing carrier-to-noise density ($C/N_0$) levels for GPS satellites tracked by the receiver, which is the ratio of received GPS signal power level to noise density. Lower $C/N_0$ levels result in increased data bit error rate when extracting navigation data from GPS signals, and hence, increased carrier and code tracking loop jitter. This, in turn, results in more noisy range measurements and thus, less precise positioning (DOD, 2001; USACE, 2003; Kaplan & Hegarty, 2006; Petovello, 2009). The increase in probable error values was also observed during GPS jamming tests conducted via field evaluations and GPS simulation.

For all the readings, the highest probable errors occur at the minimum spoofing power levels. After spoofing takes place, the probable errors reduce to levels that are lower as compared to prior to transmission of the spoofing signal. This occurs as at this point, the spoofing signal power levels are relatively large, resulting in high $C/N_0$ levels and hence, improved accuracy.

Varying probable error patterns are observed for the each of the readings. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over time, resulting in GPS accuracy being time dependent (DOD, 2001; USACE, 2003; Kaplan & Hegarty, 2006; Huihui *et al.*, 2008; Dinesh *et al.*, 2010b). Variation in other GNSS error parameters could have also resulted in the variation of probable error patterns.

Varying probable error patterns are observed for the each of the readings. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over time, resulting in GPS accuracy being time dependent (DOD, 2001; USACE, 2003; Kaplan & Hegarty, 2006; Huihui *et al.*, 2008; Dinesh *et al.*, 2010b). Variation in other GNSS error parameters could have also resulted in the variation of probable error patterns.



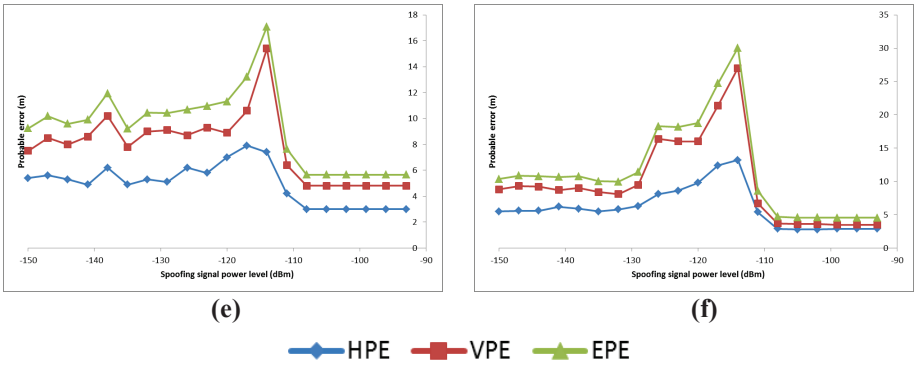
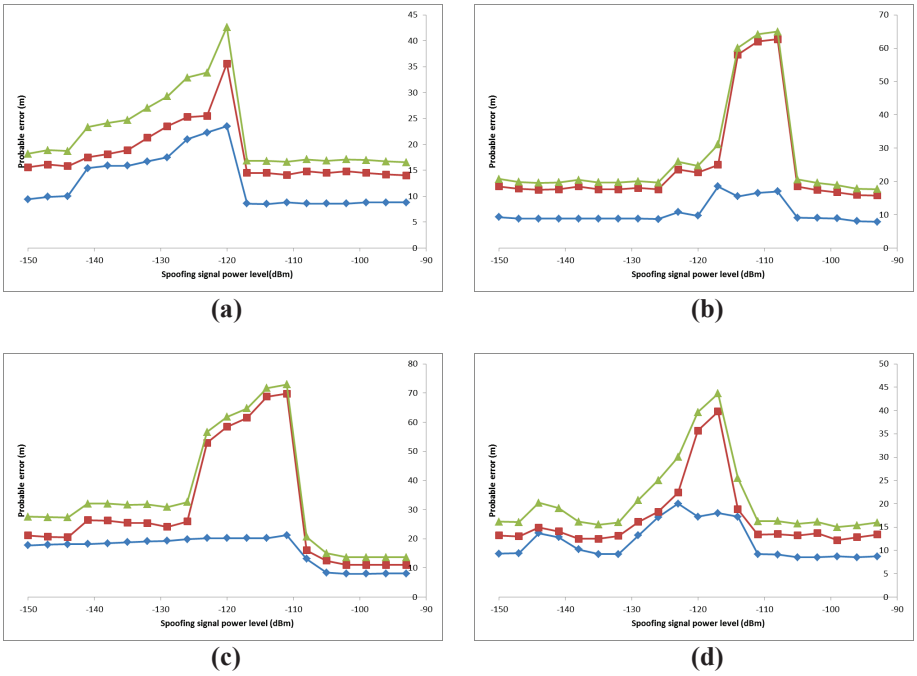



**Figure 3: Recorded probable error values for the evaluated GPS receiver: (a) Reading 1 (b) Reading 2 (c) Reading 3 (d) Reading 4 (e) Reading 5 (f) Reading 6.**



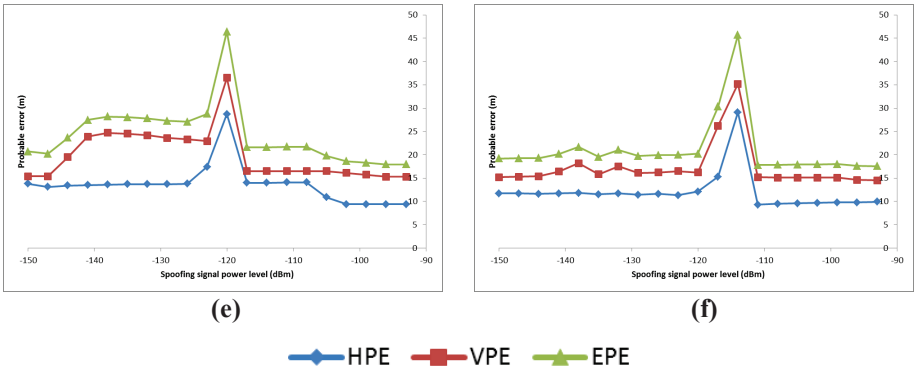



**Figure 4: Recorded probable error values for the reference GPS receiver: (a) Reading 1 (b) Reading 2 (c) Reading 3 (d) Reading 4 (e) Reading 5 (f) Reading 6.**

For all the readings taken, the evaluated GPS receiver recorded lower probable error values as compared to the reference GPS receiver. This occurred as the evaluated GPS receiver has higher receiver sensitivity, and hence, is able to obtain lower PDOP values. In addition, it has lower receiver noise, reducing the value of its user equivalent ranging error (UERE), which is the total expected magnitude of position errors due to measurement uncertainties from the various error components for a particular receiver.

It is observed that for all the readings, the values of VPE are larger than HPE, as GPS receivers can only track satellites above the horizon, resulting in GPS height solution being less precise than the horizontal solution (DOD, 2001; USACE, 2003; Kaplan & Hegarty, 2006; Huihui et al., 2008; Dinesh et al., 2010b). The difference between VPE and HPE values is significantly larger for the reference GPS receiver as compared to the evaluated GPS receiver. The reference GPS receiver, having lower receiver sensitivity, has much better horizontal component accuracy as compared to the vertical component. For the GPS evaluated receiver, with higher receiver sensitivity, while the horizontal component accuracy is still larger, the difference with the vertical component is much smaller. During the spoofing tests, at some points, depending on GPS coverage, the horizontal component accuracy became lower than the vertical component due to significant reduction of $C/N_0$ levels for overhead satellites as compared to satellites above the horizon.

On the whole, this test has demonstrated the disadvantages of field GNSS evaluations. Without the ability to control the various GNSS error parameters, it is difficult to effectively study the effect of any particular error parameter, in the case of this test, interference, on GNSS accuracy. This highlights the importance of conducting such tests in a controlled environment, using a GNSS simulator the

source of genuine GNSS signals as opposed to live GNSS signals. This would allow the tests to be conducted under repeatable user-controlled conditions.

## 4. CONCLUSION

This study has highlighted the effect of simplistic spoofing attacks on GPS performance. Varying minimum spoofing signal power levels, times between position fix lost and spoofing, and probable error patterns are observed for different dates and times. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over time, resulting in GPS performance being time dependent. Variation in other GNSS error parameters, including ionospheric and tropospheric delays, satellite clock, ephemeris and multipath errors, and unintentional signal interferences and obstructions, could have also resulted in the variation of GPS performance. As the spoofing signal power level is increased, probable error values increase due to decreasing $C/N_0$levels for GPS satellites tracked by the receiver. For all the readings, the highest probable errors occur at the minimum spoofing power levels. After spoofing takes place, the probable errors reduce to levels that are lower as compared to prior to transmission of the spoofing signal. This occurs as at this point, the spoofing signal power level is relatively large, resulting in high $C/N_0$ level and hence, improved accuracy. Due to the disadvantages of field GNSS evaluations, this study should be repeated in a controlled environment, using a GNSS simulator as the source of genuine GNSS signals as opposed to live GNSS signals, to ensure that more accurate results are obtained.

## ACKNOWLEDGEMENT

The authors are grateful to Dr. Todd Humphreys, Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, for his suggestions that have helped strengthen this manuscript.

## REFERENCES

Advantest (2009). *U3741/3751 Spectrum Analyzers*. Advantest Corporation, Chiyoda-ku, Tokyo

Aeroflex (2010). *Avionics GPSG-1000 GPS / Galileo Portable Positional Simulator*. Aeroflex Inc., Plainview, New York.

Cavaleri, A., Motella, B., Pini, M. & Fantino, M. (2010). Detection of spoofed GPS signals at code and carrier tracking level. *2010 5th ESA Workshop on Satellite Navigation Technologies and European*

*Workshop on GNSS Signals and Signal Processing (NAVITEC)*, 8-10 December 2010, Noordwijk.

CNET (2004). GPSDiag 1.0. Available online at: http://download.cnet.com/GPSDiag/3000-2130_4-4951103.html (Last access date: 31 January 2010).

Daneshmand, S., Jafarnia-Jahromi, A., Broumandan, A. & Lachapelle, G. (2011). *Low-Complexity Spoofing Mitigation*. Available online at: http://www.gpsworld.com/gnss-system/signal-processing/low-complexity-spoofing-mitigation-12366?utm_source=GPS&utm_medium=email&utm_campaign=Defense-PNT_12_14_2011&utm_content=low-complexity-spoofing-mitigation-12366 (Last access date: 16 December 2011).

Department of Defence (DOD) (2001). *Global Positioning System Standard Positioning Service Performance Standard*. Command, Control, Communications, And Intelligence, Department of Defence (DOD), Washington D.C.

Dinesh, S. (2009). Vulnerabilities of civilian Global Navigation Satellite Systems (GNSS) signals: A review. *Defence S&T Tech. Bull.*, **2**: 100-114.

Dinesh, S. (2011). Vulnerabilities of Global Navigation Satellite Systems (GNSS) signals to jamming and spoofing. *GIS Day 2011*, 17 March 2011, Universiti Putra Malaysia (UPM), Serdang, Selangor.

Dinesh , S., Wan Mustafa, W.H., Mohd Faudzi., M., Kamarulzaman, M., Nor Irza Shakhira, B., Siti Robiah, A., Norhayaty, Z., Aliah, I., Lim, B.T., Arumugam, P., Zainal Fitry, M.A., Mohd. Rizal, A.K., Azlina, B. & Mohd. Hasrol, H.M.Y. (2009). Evaluation of the effect of radio frequency interference (RFI) on Global Positioning System (GPS) receivers. *Defence S&T Tech. Bull.*, **2**: 115-129.

Dinesh , S., Wan Mustafa, W.H., Mohd Faudzi., M., Kamarulzaman, M., Hasniza, H., Nor Irza Shakhira, B., Siti Robiah, A., Shalini, S., Jamilah, J., Aliah, I., Lim, B.T., Zainal Fitry, M.A., Mohd. Rizal, A.K., Azlina, B. & Mohd. Hasrol, H.M.Y. (2010a). Evaluation of power levels required by interference signals at various distances to jam the Global Positioning System (GPS) L1 coarse acquisition (C/A) signal. *Defence S&T Tech. Bull.*, **3**: 14-28.

Dinesh , S., Wan Mustafa, W.H., Mohd Faudzi., M., Kamarulzaman, M., Hasniza, H., Nor Irza Shakhira, B., Siti Robiah, A., Shalini, S., Jamilah, J., Aliah, I., Lim, B.T., Zainal Fitry, M.A., Mohd. Rizal, A.K., Azlina, B. & Mohd. Hasrol, H.M.Y. (2010b). Evaluation of the effect of radio frequency interference (RFI) on Global Positioning System (GPS) accuracy. *Defence S&T Tech. Bull.*, **3**: 100-118.

Dinesh, S., Mohd Faudzi., M., Rafidah, M., Nor Irza Shakhira, B., Siti Robiah, A., Shalini, S., Jamilah, J., Aliah, I., Lim, B.T., Zainal Fitry, M.A., Mohd. Rizal, A.K., & Mohd Hasrol, H.M.Y. (2012a). Evaluation of the effect of

radio frequency interference (RFI) on Global Positioning System (GPS) receivers via GPS simulation. *ASM Sci. J.*, Under review.

Dinesh, S., Mohd Faudzi., M. & Zainal Fitry, M.A. (2012b). Evaluation of the effect of radio frequency interference (RFI) on Global Positioning System (GPS) accuracy via GPS simulation. *Defence Sci. J.*, **62**: 338-347..

Garmin (2004). *GPSmap 60CS Owner's Manual*. Garmin International Inc., Olathe, Kansas.

Garmin (2007). *GPSmap 60CSx Owner's Manual*. Garmin International Inc., Olathe, Kansas.

GPS Source (2006). *L1P GPS Antenna. GPS Source*. GPS Source Inc., Pueblo West, Colarado.

GPS Source (2007). *A11XLV Digital Variable Gain GPS Amplifier*. GPS Source Inc., Pueblo West, Colarado.

Hanlon, B.O., Ledvina, B., Psiaki, M.L., Kintner. P.M. & Humphreys, T.E. (2009). *Assessing the Spoofing Threat*. Available online at:

http://www.gpsworld.com/defence/security-surveillance/assessing-spoofing-threat-3171?page_id=1 (Last access date: 4 November 2009).

Huihui, W., Xingqun, Z. & Yanhua, Z. (2008). Geometric dilution of precision for GPS single-point positioning based on four satellites. *J. Syst. Eng. Electr.*, **19**: 1058-1063.

Humphreys, T.E., Ledvina, B.M., Psiaki, M.L., & Kintner, J. (2008). Assessing the spoofing threat: Development of a portable GPS civilian spoofer. *ION GNSS 2009*, 16-19 September 2008, Savannah International Convention Center, Savannah, Georgia.

Humphreys, T.E., Psiaki, M.L. & Kintner, P.M. (2009). *GPS Spoofing Threat*. Available online at: http://www.telecomasia.net/article.php?id_article=12288&page=4 (Last access date: 4 November 2009).

Johnston, R.G. & Warner, J.S. (2004). Think GPS offers high security? Think again! *Business Contingency Planning Conference*, 23-27 May 2004, Las Vegas, Nevada.

Jones, M. (2011). The civilan battlefield: Protecting GNSS receivers from interference and jamming. *Inside GNSS,* **6**: 40-49.

Kaplan, E.D. & Hegarty, C.J. (2006). *Understanding GPS: Principles and Applications*. Artech House, Norwood, Massachusetts.

Key, E.L. (1995). *Techniques to Counter GPS Spoofing*. Internal memorandum, MITRE Corporation, 17 February 1995.

Last, D. (2010). GNSS: The present imperfect. *Inside GNSS*, **5**: 60-64.

Ledvina, B., Montgomery, P., & Humphreys, T. (2009). A multi-antenna defense: Receiver-autonomous GPS spoofing detection. *Inside GNSS*, **4**: 40-46.

Montgomery, P., Humphreys, T.E. & Ledvina, B.M. (2009). A multi-antenna defence receiver-autonomous GPS spoofing detection. *Inside GNSS*, **4**: 40-46.

Motella, B., Pini, M., Fantino, M., Mulassano, P., Nicola, M., Fortuny-Guasch, J., Wildemeersch, M. & Symeonidis, D. (2010). Performance assessment of low cost GPS receivers under civilian spoofing attacks. *2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)*, 8-10 December 2010, Noordwijk.

Nicola, M., Musumeci, M., Pini, M., Fantino, M. & Mulassano, P. (2010). Design of a GNSS spoofing devicebased on a GPS/Galileo software receiver for the development of robust countermeasures. *European Navigation Conference 2010 (ENC 2010)*, 19-21 October 2010, Braunchweig, Germany.

Nielsen, J., Broumandan & Lachappelle, G. (2010). *Spoofing Detection and Mitigation*. Available online at:

http://www.gpsworld.com/gnss-system/receiver-design/spoofing-detection-and-mitigation-10456?page_id=2 (Last access date: 31 January 2011).

Papadimitratos, P. & Jovanovic, A. (2008). Protection and fundamental vulnerability of GNSS. *International Workshop on Satellite and Space Communications 2008 (IWSSC'08)*, 1-3 October 2008, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France.

Petovello, M. (2009). Carrier-to-noise density and AI for INS / GPS integration. *Inside GNSS*, **4**: 20-29.

Pozzobon, O., Wullems, C., & Detratti, M. (2011). *Tamper Resistance: Security Considerations for GNSS Receivers*. Available online at: http://www.gpsworld.com/transportation/tamper-resistance-11403 (Last access date: 15 April 2011).

Rogers, C. (1991). Development of a low cost PC controlled GPS satellite signal simulator. *Proceedings of the 15th Biennial Guidance Test Symposium*, Holloman AFB, New Mexico.

Royal Academy of Engineering (RAE) (2011). *Global Navigation Space Systems: Reliance and Vulnerabilities*. Royal Academy of Engineering, London.

Schue, C. (2012). The challenges of realizing a global navigation capability. *ION International Technical Meeting (ITM) 2012*, 30 January - 1 February 2012, Newport Beach, California.

Schwartz, N. (2010). The United States as an aerospace nation: Challenges and opportunities. *Tuft University Institute for Foreign Policy Analysis (IFPA) Fletcher Conference on National Security Strategy and Policy*, 20-21

January 2010, The Ronald Reagan Building and International Trade Center, Washington, D.C.

Scott, L.S. (2012). Spoofs, proofs & jamming: Towards a sound national policy for civil location and time assurance. *Inside GNSS*, **7**: 42-53.

Sherman, L., Dennis, A., Eklöf, F.M., I.O. & Borowski, H. (2012). *Detecting False Signals with Automatic Gain Control*. Available online at: http://www.gpsworld.com/government/detecting-false-signals-automatic-gain-control-12804 (Last access date: 15 May 2012).

Shepard, D.P. & Humphreys, T.E. (2011). Characterization of receiver response to spoofing attacks. *ION GNSS 2011*, 19-23 September 2011, Portland.

Shepard, D.P., Bhatti, J.A. & Humphreys, T.E. (2012). Evaluation of smart grid and civilian UAV vulnerability to GPS spoofing attacks. *ION GNSS 2012*, 17-21 September 2012, Nashville Convention Center, Nashville, Tennessee.

Tippenhauer, N.O., Pöpper, C., Rasmussen, K.B. & Čapkun, S. (2011). On the requirements for successful GPS spoofing attacks. *ACM Conference on Computer and Communications Security (CCS)*, 17-21 October 2011, Chicago, Illinois.

Trimble (2012). *Trimble's Planning Software*. Available online at:

http://www.trimble.com/planningsoftware.shtml (Last access date: 10 July 2012).

US Army Corps of Engineers (USACE) (2003). NAVSTAR Global Positioning System Surveying. Engineer Manual, EM 1110-1-1003, US Army Corps of Engineers (USACE), Washington D.C.

Volpe (2001). Vulnerability Assessment of the Transport Infrastructure Relying on the Global Positioning System. John A. Volpe National Transportation Systems Center, Department of Transport, Washington D.C.

Warner, J.S. & Johnston, R.G. (2002). A simple demonstration that the global positioning system (GPS) is vulnerable to spoofing. *J. Secur. Admin.*, **25**: 19-28.

Wen, H., Huang, P.Y.R., Dyer, J., Archinal, A. & Fagan, J. (2005). Countermeasures for GPS signal spoofing. *18th International Technical Meeting of the Satellite Division of the Institute of Navigation ION GNSS 2005*, 13-16 September 2005, Long Beach Convention Center, Long Beach, California.

Wesson, K., Shephard, D. & Humphreys, T.E.(2012). *Straight Talk on Anti-Spoofing*. Available online at:

http://www.gpsworld.com/GNSS%20System/Signal%20Processing/straight-talk-anti-spoofing-12471 (Last access date: 10 January 2012).

# PERFORMANCE ANALYSIS OF BIO-INSPIRED ROUTING PROTOCOLS BASED ON RANDOM WAYPOINT MOBILITY MODEL

Vaibhav Godbole*

Department of Information Technology, Fr. Conceicao Rodrigues College of Engineering, India

*E-mail: vai.godbole@gmail.com

## ABSTRACT

*A mobile ad hoc network (MANET) is a non-centralised, multihop, wireless network that lacks a common infrastructure and hence it needs self-organisation. The biggest challenge in MANETs is to find a path between communicating nodes, which is the MANET routing problem. Biology-inspired techniques such as ant colony optimisation (ACO) which have proven to be very adaptable in other problem domains, have been applied to the MANET routing problem as it forms a good fit to the problem. The general characteristics of these biological systems, which include their capability for self-organisation, self-healing and local decision making, make them suitable for routing in MANETs. In this paper, we discuss a few ACO based protocols, namely AntNet, hybrid ACO (AntHocNet), ACO based routing algorithm (ARA), imProved ant colony optimisation routing algorithm for mobile ad hoc NETworks (PACONET), ACO based on demand distance vector (Ant-AODV) and ACO based dynamic source routing (Ant-DSR), and determine their performance in terms of quality of service (QoS) parameters, such as end-to-end delay and packet delivery ratio, using Network Simulator 2 (NS2). We also compare them with well known protocols, ad hoc on demand distance vector (AODV) and dynamic source routing (DSR), based on the random waypoint mobility model. The simulation results show how this biology-inspired approach helps in improving QoS parameters.*

**Keywords:** *Mobile ad hoc network (MANET); ant colony optimisation (ACO); random waypoint model; Network Simulator 2 (NS2); quality of service (QoS) parameters.*

## 1. INTRODUCTION

Ad-hoc networks can be classified in three categories based on their applications; mobile ad-hoc networks (MANETs), wireless mesh networks (WMNs) and

wireless sensor networks (WSN) (Mann and Mazhar, 2011). MANET is a collection of mobile nodes with a wireless network interface which forms a temporary network without the aid of any fixed infrastructure or centralised administration. Nodes within each other's transmission ranges can communicate directly, but nodes outside each other's range have to rely on other nodes to transmit the messages. Figure 1 shows a basic ad-hoc network. In this network, packet transmission from source to destination takes place without a base station.

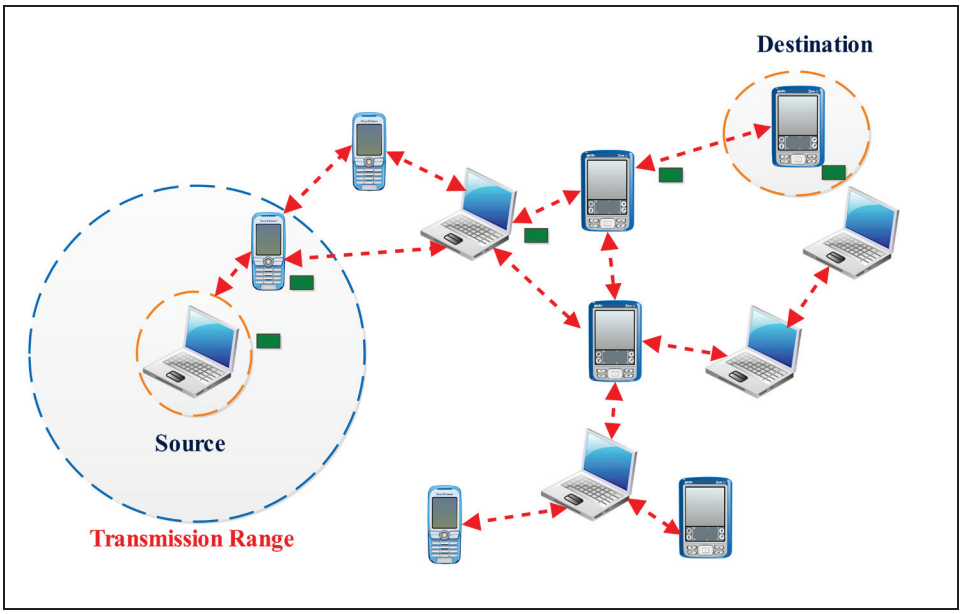



**Figure 1: A basic ad-hoc network.**
**(Source: Villalba *et al.* (2011))**

Research interest in MANETs has been growing in past few years, with the design of MANET routing protocols receiving significant attention. One of the reasons for this is that routing in MANETs is a particularly challenging task due to the fact that the topology of the network changes constantly, and paths which were initially efficient can quickly become inefficient or even infeasible. Moreover, control of information flow in the network is very restricted. This is because the bandwidth of the wireless medium is very limited, and as the medium is shared, nodes can only send or receive data if no other node is sending in their radio neighbourhood (Abolhasan *et al.*, 2004).

Basically, routing protocols in MANETs are classified into three categories; proactive, reactive and hybrid (Abolhasan *et al.*, 2004). Proactive routing protocols often need to exchange control packets among mobile nodes and continuously update their routing tables. Each node must maintain the state of the network in real time. This causes high overhead congestion of the network, which requires a lot of memory. The advantage of proactive protocols is that

nodes have correct and updated information. Hence, when a path is required, it can be found directly in the memory and links can be established quickly. These protocols are intended to reduce broadcasting frequency while maintaining correct information for the routing table. Reactive routing protocols only seek a route to the destination when it is needed. The advantage of these protocols is that the routing tables located in the memory are not continuously updated. On the other hand, they have the disadvantage that they cannot establish connections in real time. The aim of these protocols is to save time in the route discovery process, since the reactive protocol is designed to reduce the latency which is critical in this kind of protocols. It also aims to avoid the maintenance of routes to prevent long delay (Singla & Kakkar, 2010).

It is therefore important to design protocols that are adaptive, robust and self-healing. Moreover, they should work in a localised way, due to the lack of central control or infrastructure in the network. Nature's self-organising systems, such as insect societies, termite hills, bee colonies, bird flocks and fish schools, provide precisely these features and hence have been a source of inspiration for the design of many routing algorithms for MANETs (Abdel-Moniem *et al.*, 2010). In this paper we discuss few ant colony optimisation (ACO) based routing protocols for MANETs, namely AntNet, hybrid ACO (AntHocNet), ACO based routing algorithm (ARA), improved ant colony optimisation routing algorithm for mobile ad hoc NETworks (PACONET), ACO based on demand distance vector (Ant-AODV) and ACO based dynamic source routing (Ant-DSR). We choose these protocols for analysis because all these protocols are based on forward ants (FANT) and backward ants (BANT) principle We compare these protocols based on quality of service (QoS) parameters such as end-to-end delay and packet delivery ratio, with conventional protocols such as ad hoc on-demand distance vector (AODV) and dynamic source routing (DSR), based on the random waypoint mobility model.

The random waypoint mobility model includes pause times between changes in direction and/or speed. A mobile node (MN) begins by staying in one location for a certain period of time (i.e., a pause time). Once this time expires, the MN chooses a random destination in the simulation area and a speed that is uniformly distributed between the minimum speed and maximum speed. The MN then travels toward the newly chosen destination at the selected speed. Upon arrival, the MN pauses for a specified time period before starting the process again (Camp & Davies, 2002). Figure 2 shows the travelling pattern of a MN using random waypoint mobility model.

## 2. ANT COLONY OPTIMISATION (ACO)

### 2.1 Similarities between Ad Hoc Networks and Ants

There are lots of similarities between ad hoc networks and ants, such as shown in Table 1. Ant based routing algorithms exhibit a number of desirable properties for ad hoc networks. The foraging behaviour of ants and bees, and the hill building behaviour of termites have inspired researchers in developing efficient routing algorithm for ad hoc networks (Gupta *et al.*, 2012).

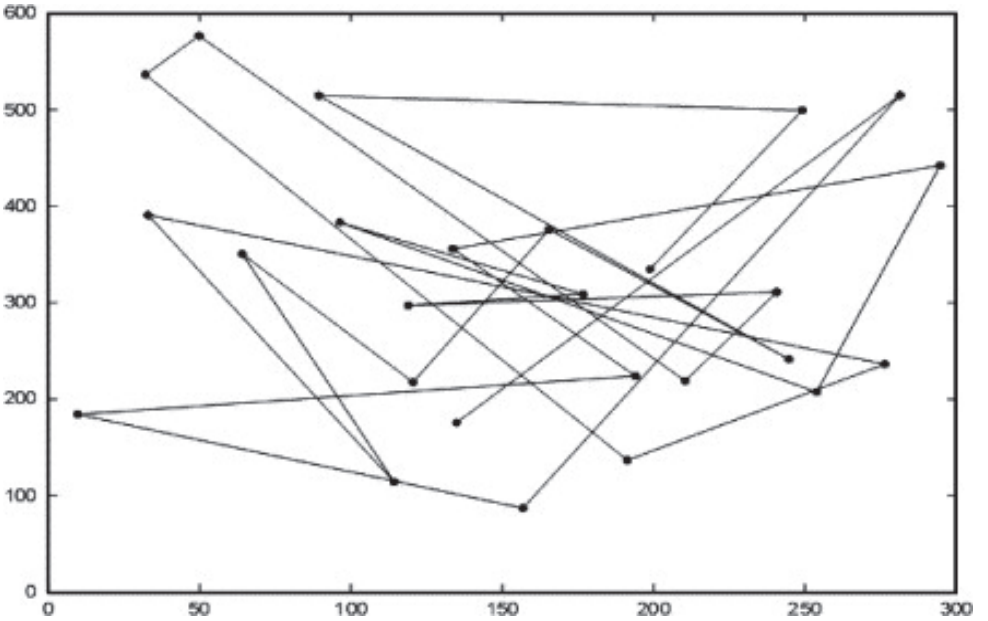



**Figure 2: Traveling pattern of a MN using the random way point mobility model. (Source: Camp & Davis, 2002)**

**Table 1: Comparison between ad hoc networks and ants. (Source: Gupta *et al*., 2012)**

| Parameters | Ad-hoc networks | Ants |
|---|---|---|
| Physical structure | Unstructured, dynamic, distributed | Unstructured, dynamic, distributed |
| Origin of route | Route requests are sent from source to get local information | Pheromones are used to build new routes |
| Multipath support | Single path, partially multipath | Provide multipath |
| Basic system | Self-configuring, Self-organising | Self-configuring, Self-organising |

## 2.2 Ants in Nature

The main source of inspiration behind ACO is a behaviour that is displayed by certain species of ants in nature during foraging. It has been observed that ants are able to find the shortest path between their nest and a food source. The only way that this difficult task can be realised is through the cooperation between the individuals in the colony (Caro & Dorigo, 1998).

The key behind the colony level shortest path behaviour is the use of pheromone. This is a volatile chemical substance that is secreted by the ants in order to influence the behaviour other ants and of it. Pheromone is not only used by ants to find shortest paths, but is in general is an important tool that is used by many different species of ants (Caro & Dorigo, 1998).

Ants moving between their nest and a food source leave a trail of pheromone behind, and they also preferably go in the direction of high intensities of pheromone. We use the example situation depicted in Figure 3 to explain how this simple behaviour leads to the discovery of shortest paths. In our example, there are two possible paths between the ant nest and the food source, one of which is considerably shorter than the other. The first ants leaving the nest have no information available. They therefore choose their movements randomly. This leads to approximately 50% of the ants choosing the short path and 50% choosing the long path (Jha *et al*., 2011). All moving ants leave a trail of pheromone behind. The ants going over the short path reach the destination earlier than those going over the long path. Moreover, they can return faster. This leads temporarily to a higher pheromone concentration on the shortest path. Subsequent ants leaving the nest are attracted by this higher intensity, and go therefore preferably also over the shortest path. As this process continues, the majority of the ants eventually concentrate on the shortest path. However, it should be pointed out that the behaviour of the ants is never deterministic and hence, there will always remain a minority of ants that explore the longer path (Marwaha *et al.*, 2002).

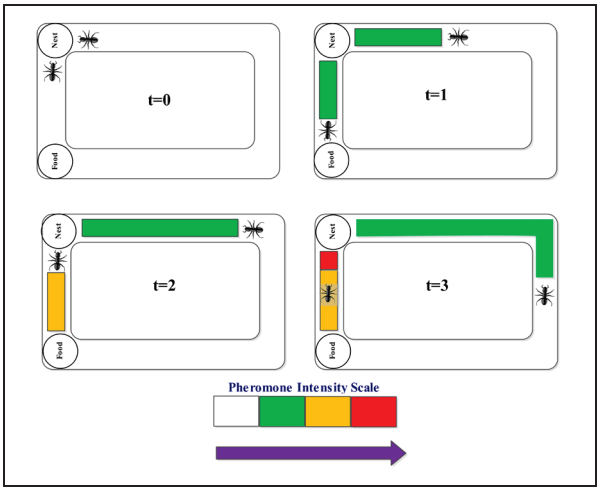



**Figure 3: The shortest path mechanism used by ants. The different colours indicate increasing levels of pheromone intensity. The scenario is depicted in successive time steps *t*.**
**(Source: Baluja & Davies, 1998)**

## 2.3 Routing in Ant Systems

The core of any network control system is routing which strongly affects the overall network performance. Routing deals with the problem of defining the path to forward incoming data traffic such that the overall network performance is maximised. At each node, data is forwarded according to the content of the routing table, maintains the information of source address, destination address, ant rip times etc. In this sense, a routing system can be seen as a distributed decision system (Jha *et al*., 2011). A variety of different classes of specific routing can be defined according to the different characteristics of processing, transmission components, traffic pattern and type of performance (Baluja & Davies 1998).

The following set of core properties of ACO characteristics for routing problems (Baluja & Davies 1998).

- provide traffic-adaptive and multipath routing
- rely on both passive and active information monitoring and gathering
- make use of stochastic components
- do not allow local estimates to have global impact
- set up paths in a less selfish way than in pure shortest path schemes favouring load balancing

These are all characteristics that directly result from the application of ACO's design guidelines, and in particular, from the use of controlled random

experiments (the ants) that are repeatedly generated in order to actively gather useful non-local information about the characteristics of the solution set (i.e., the set of paths connecting all pairs of source-destination nodes in the routing case) (Baluja & Davies 1998).

## 3. ACO ALGORITHMS

In the nature, ants lay pheromone and so they produce pheromone trails between the nest and a food source. On a computer, the pheromone has been replaced by artificial stigmergy, the probabilities in the routing tables. To compute and update the probabilities, intelligent agents are introduced to replace the ants. There exist two kinds of agents, forward and backward agents. All forward and backward agents have the same structure. The agents move inside the network by hopping at every time step from a node to the next node along the existing links. The agents communicate with each other in an indirect way by concurrently reading and writing the routing tables on their way (Villalba & Orozco, 2010). In this section, we discuss based routing protocols for MANETs, namely AntNet, AntHocNet, ARA, PACONET, Ant-AODV and Ant-DSR.

### 3.1 AntNet

AntNet is a direct extension of the simple ACO algorithm (Di Caro, 1998). AntNet is even closer to the real ant's behaviour that inspired the development of the ACO meta-heuristic than the original ACO algorithms (Jha *et al*. 2011).

AntNet is conveniently described in terms of two sets of artificial ants, called as forward ant (FANT) and backward ant (BANT). Ants in each set possess the same structure, but they are situated differently in the environment; that is, they can sense different inputs and they can produce different, independent outputs. Ants communicate in an indirect way, according to the stigmergy paradigm, through the information they concurrently read and write on the network nodes they visit (Di Caro, 1998).

At regular intervals$\Delta$ $t$ from every network node $s$, a FANT $F_{s\rightarrow d}$ is launched towards a destination node $d$ to discover a feasible, low-cost path to that node and to investigate the load status of the network along the path. FANTs share the same queues as data packets, so that they experience the same traffic load. Destinations are locally selected according to the data traffic patterns generated by the local workload: if $f_{sd}$ is a measure (in bits or in the number of packets) of the data flow $s \rightarrow d$, then the probability of creating at node $s$ a FANT with node $d$ as destination is given by (Villalba & Orozco, 2010):

$$P_{sd} = \frac{f_{sd}}{\sum_{i=1}^{n} f_{si}} \qquad (1)$$

The ant builds a path using the following steps:

1. At each node $i$, each FANT headed toward a destination {\it $d$} selects the node $j$ to move to, choosing among the neighbours it did not already visit, or over all the neighbours in case all of them had previously been visited. The neighbour $j$ is selected with a probability $P_{ijd}$ computed as the normalised sum of the pheromone $\tau_{ijd}$ with a heuristic value $n_{ij}$ taking into account the state (the length) of the $j^{th}$ rank link queue of the current node $i$:

$$P_{ijd} = \frac{\tau_{ijd} + \alpha \eta_{ij}}{1 + \alpha(|\mathcal{N}_i| - 1)} \tag{2}$$

   The heuristic value $n_{ij}$ is a [0, 1] normalised values function of the length $q_{ij}$ (in bits waiting to be sent) of the queue on the link connecting the node $i$ with its neighbour $j$:

$$\eta_{ij} = 1 - \frac{q_{ij}}{\sum_{l=1}^{|\mathcal{N}_i|} q_{il}} \tag{3}$$

The value of $\alpha$ weighs the importance of the heuristic value with respect to the pheromone values stored in the pheromone matrix $T$. The value $h_{ij}$ reflects the instantaneous state of the node's queues and, assuming that the queue's consuming process is almost stationary or slowly varying, $h_{ij}$ gives a quantitative measure associated with the queue waiting time (Caro & Dorigo, 1998).

2. When the destination node $d$ is reached, the agent $F_{s \to d}$ generates another agent, BANT $B_{d \to s}$, transfers to all of its memory, and is deleted. A FANT is also deleted of its lifetime and becomes greater than a value *max_life* before it reaches its destination node, where *max_life* is a parameter of the algorithm.

3. The BANT takes the same path as that of its corresponding FANT, but in the opposite direction. BANTs do not share the same link queues as data packets; they use higher-priority queues reserved for routing packets, because their task is to quickly propagate to the pheromone matrices the information accumulated by the FANTs (Sujatha & Harigovindan, 2010).

4. The re-enforcement factor $r$ is defined as the ratio of travel time of an ant at a specific node to the travel time of all ants at that node. The value of $r$ is such that $0<r<1$. This factor is pre-defined in the AntNet algorithm.

### 3.2 Hybrid ACO (AntHocNet)

AntHocNet (Di Caro *et al*., 2005) combines the typical path sampling behaviour of ACO algorithms with a pheromone bootstrapping mechanism. AntHocNet is a hybrid algorithm. It is reactive in the sense that a node only starts gathering routing information for a specific destination when a local traffic session needs to communicate with the destination and no routing information is available. It is proactive because as soon as the communication starts, and for the entire duration of the communication, the nodes proactively keep the routing information related to the on-going flow up-to-date with network changes for both topology and traffic. The algorithm tries to find paths characterised by minimal number of hops, low congestion and good signal quality between adjacent nodes.

Nodes in AntHocNet forward data stochastically. When a node has multiple next hops for the destination d of the data, it randomly selects one of them with probability $P_{nd}$. $P\$_{nd}$, which is calculated as follows:

$$P_{nd} = \frac{(\tau_{nd}^{i})^{\beta}}{\sum_{j \in N_d^i} (T_{jd}^{i})^{\beta}} \quad \beta >> 1 \tag{4}$$

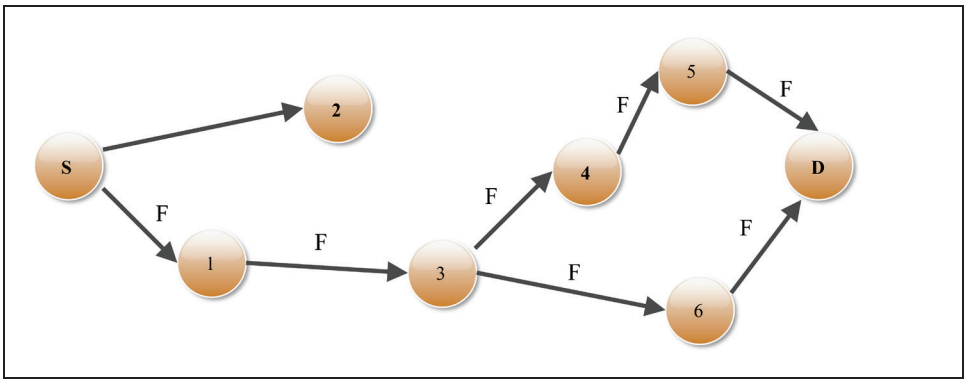
where $N^i_d$ is the set of neighbours of $i$ over which a path to $d$ is known, and $\beta$ is a parameter value which can control the exploratory behaviour of the ants.

The probabilistic routing strategy leads to data load spreading according to the estimated quality of the paths. When a path is clearly worse than others, it will be avoided, and its congestion will be relieved. Other paths will get more traffic, leading to higher congestion, which will make their end-to-end delay increase. By continuously adapting the data traffic, the nodes try to spread (Lin & Shao, 2010).

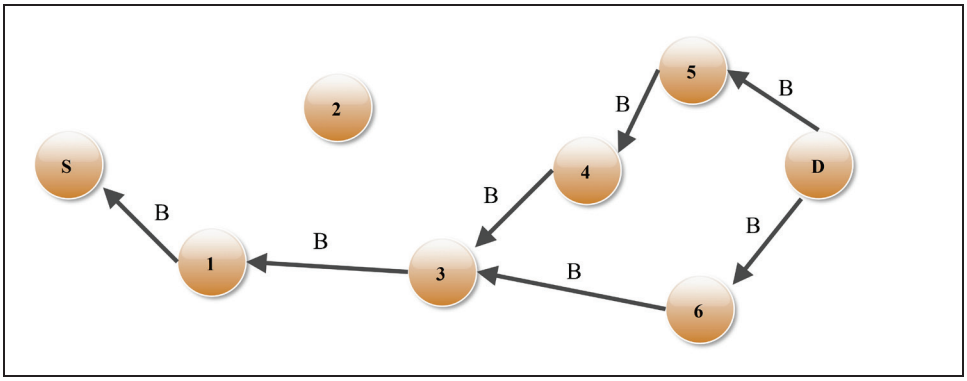
A node which receives multiple copies of the same ant only accepts the first and discards the others. When a FANT arrives at destination, it goes backward, updates the pheromone tables at the nodes, indicating a path between s and d, and triggers the sending of data packets from the traffic session. In this way, only one path is set up initially. During the course of the communication session, additional paths are added and / or removed via a proactive path maintenance and exploration mechanism. This is implemented through a combination of ant path sampling and slow-rate pheromone diffusion and bootstrapping, which mimics pheromone diffusion in nature. This way, promising pheromone is checked out, and if the associated path is there and has the expected good quality, it can be turned into a regular path available for data (Di Caro *et al.*, 2005).

### 3.3 ACO Based Routing Algorithm (ARA)

ARA (Gunes & Sorges, 2002) is a purely reactive MANET routing algorithm. It does not use any HELLO packets to explicitly find its neighbours. . HELLO packets are sent by the routers to compute the time delay to send and receive datagrams to and from its neighbors. A HELLO packet also consists of clock and timestamp information. When a packet arrives at a node, the node checks it to see if routing information is available for destination *d* in its routing table. From Figure 4, we can see that route discovery is done either by the FANT flood technique or FANT forward technique. In the FANT flooding scheme, when a FANT arrives to any intermediate node, the FANT is flooded to all its neighbours. If found, it forwards the packet over that node; if not, it broadcasts a FANT to find a path to the destination. By introducing a maximum hop count on the FANT, flooding can be reduced. In the FANT forwarding scheme, when a FANT reaches an intermediate node, the node checks its routing table to see whether it has a route to the destination over any of its neighbours. If such a neighbour is found, the FANT is forwarded to only that neighbour; else, it is flooded to all its neighbours as in the flood scheme. In ARA, a route is indicated by a positive pheromone value in the node's pheromone table over any of its neighbours to the FANT destination. When the ant reaches the destination it is sent back along the path it came, as a backward ant as shown in Figure 4(b). All the ants that reach the destination are sent back along their path. Nodes modify their routing table information when a backward ant is seen according to number of hops the ant has taken. When a route is found, the packet is forwarded over the next hop stochastically.




**(a)**




**(b)**

**Figure 4: Route discovery phase: (a) A FANT *F* is send from the sender *S* toward the destination node *D*. The forward ant is relayed by other nodes, which initialise their routing table and the pheromone values. (b) The BANT *B* has the same task as *F*. It is sent by *D* towards *S*.**
**(Source: Gunes *et al.*, 2002)**

### 3.4 imProved ant colony optimisation routing algorithm for mobile ad hoc NETworks (PACONET)

PACONET (Osagie & Thulasiram, 2008) is a routing protocol for mobile ad-hoc networks inspired by the foraging behaviour of ants. It uses the principles of ACO routing to develop a suitable problem solution.

The PACONET protocol is based on the parallel ACO algorithm. The availability of parallel architectures at low cost has widened the interest for the parallelisation of algorithms and metaheuristics. When developing parallel genetic algorithms and parallel ACO algorithms, it is common to adopt the strategy of information exchange that plays a major role in the algorithms. Solutions, pheromone matrices, and parameters have been tested as the object of such an exchange (Gunes & Sorges, 2002).

The fundamental principle of parallel ant colony algorithm (Middendorf *et al.*, 2002; Chengyong Liu & Xiang, 2008) is to divide $M$ ants into $P$ ant colonies. Normally, the numbers of every ant colony are the same, viz. $N_a = M/P$. In the algorithm development, each colony is distributed to a processor, and then, the ant colony can search the best solution independently. In order to avoid local optimisation in some processor when the ant colony is doing the job, the processors should carry out the information exchange each other in the fixed condition (i.e. time interval, etc.) The FANT explores the paths of the network in a restricted broadcast manner in search of routes from a source to a destination. The BANT establishes the path information acquired by the FANT.

When a source node *S* wishes to communicate with a destination node *D* for which it has no route information, it sends out a FANT to all its neighbours in search of the destination node. When a FANT from *S* traveling to *D*, arrives at a node *v*, the FANT determines its path or next hop neighbour by looking at the node's routing table. It considers the node's neighbours by looking at the rows against the columns in the routing table to select the best path from a neighbouring node to *D* rather than the best link between itself and its neighbour.

The FANT will consider the pheromone concentration only when all neighbours in column have been visited. The purpose of this is to ensure that all possible paths are explored to find the best path towards the destination. The node with the highest pheromone is chosen as the next hop after the FANT has determined that it has not visited the node before. This is to avoid the ant travelling in cycles. The FANT maintains a list of all nodes visited on its journey to *D* for this purpose. The FANT keeps in memory the total time *T* it has travelled. When a next hop node $v_j$ is selected from $v_i$ the FANT moves to $v_j$ and updates the pheromone entry for ($v_i$,*S*) in $v_j$'*s* routing table using the following equation:

$$\delta(v_i, v_s) = \delta(v_i, v_j) + \frac{\varepsilon}{T(v_s, v_i) + w(v_i, v_j)} \tag{5}$$

where $\varepsilon$ is a user defined run time parameter; $\delta(v_i, v_j)$ and $w(v_i, v_j)$ represent the pheromone value on each edge and time period respectively for which the links are in connection. For all the other nodes in the source column, the pheromone values are decremented by the following equation:

$$\delta(v_l, v_s) = (1 - \xi)\delta(v_l, v_s), \forall l \neq i \tag{6}$$

where $\zeta$ is evaporation rate of the pheromone, which is also determined by the user.
The total time of the path just traversed is recorded as $T(v_s,v_i) + w(v_i,v_j)\}$. When the FANT reaches the destination, a corresponding BANT is created with the source of the FANT as its destination. The BANT travels towards its destination using the list of visited nodes acquired from the FANT while updating the pheromone concentration for the destination column. That is, to update an entry ($v_b$,$v_D$) for an ant at node $v_k$, travelling backwards from $v_b$ we look at the rows of $v_b$'s neighbouring nodes and column:

$$\delta(v_b, v_D) = \delta(v_b, v_D) + \frac{\varepsilon}{T'} \tag{7}$$

where *T'* is $T(v_s,v_d)$ - $T(v_s,v_k)$. The advantage of performing this update is that it makes it easy to determine the best available path reachable from a source and to find a path easily when another ant considers the source as its destination.

### 3.5 ACO Based on Demand Distance Vector (Ant-AODV)

Ant-AODV (Marwaha *et al*., 2002; Abdel-Moniem & Hedar, 2010) is a hybrid protocol that is able to provide reduced end-to-end delay and high connectivity as compared to AODV. AODV does the reactive part and an ant-based approach does the proactive one. The main goal of the ant algorithm here is to continuously create routes in the attempt to reduce the end-to-end delay and the network latency, increasing the probability of finding routes more quickly, when required. Ant-AODV's artificial pheromone model is based on the number of hops and its goal is to discover the network topology, without any other specific functions, as opposed to most ACO algorithms. Route establishment in conventional ant-based routing techniques is dependent on the ants visiting the node and providing it with routes. The nodes also have capability of launching on-demand route discovery to find routes to destinations. The use of ants with AODV increases the node connectivity (the number of destinations for which a node has unexpired routes), which in turn reduces the amount of route discoveries and also the route discovery latency. This makes the Ant-AODV hybrid routing protocol suitable for real-time data and multimedia communications. Ant-AODV uses route error (RERR) messages to inform upstream nodes of a local link failure similar to AODV. The routing table in Ant-AODV is common to both the ants and AODV. Frequent HELLO broadcasts are used to maintain a neighbour table.

### 3.6 ACO Based Dynamic Source Routing (Ant-DSR)

Ant-DSR (Fenouche & Mellouk, 2007) is a reactive protocol that implements a proactive route optimisation method through constant verification of cached routes. This approach increases the probability of a given cached route expressed through the network reality. Mobile nodes are required to maintain route caches that contain the source routes of which the mobile node is aware. Entries in the route cache are continuously updated as new routes are learnt. The protocol consists of two major phases; route discovery and route maintenance. In Ant-DSR, the FANT and BANT packets are added in the route request and reply of DSR respectively. FANTs are used to explore new paths in the network, and measure the current network state for instance by trip times, hop count or Euclidean distance travelled. BANTs serve the purpose of informing the originating node about information collected by the FANTs.

## 4. PERFORMANCE EVALUATION

### 4.1 Simulation Parameters

Network Simulator 2 (NS2) is an open-source network simulator, which is used by the researchers to analyse performance of wired and wireless networks. Due

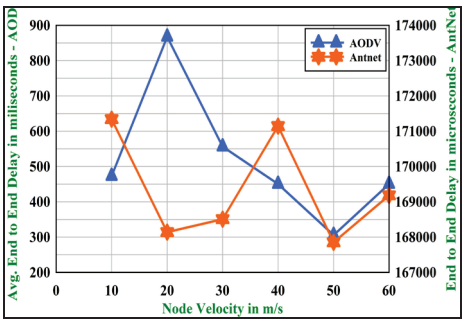
to open source nature of NS2, new protocols can be implemented by any individual. Once simulation is done, a trace file is generated. We have used awk scripts to analyse the trace files to obtain QoS parameters such as end to end delay and packet delivery ratio. Table 2 shows the simulation parameters for the ACO protocols used in this study.

**Table 2: Simulation parameters for the ACO protocols used in this study.**

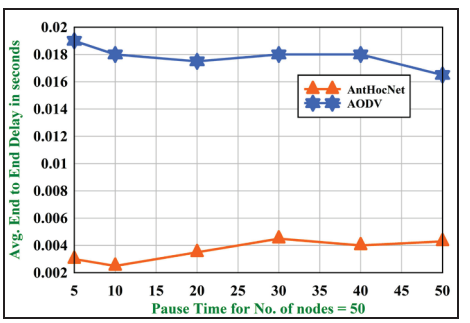
| Protocol | Parameters | | | | | |
|---|---|---|---|---|---|---|
| | No. of Nodes | Network Area ($m^{2)}$ | Simulation Time (s) | The Radio Propagation Range (m) | Packet Size | *r* |
| **AntNet** | 50 | 500 X 500 | 300 | 300 | 27 bytes. | 0.1 |
| **AntHocNet** | 50 | 500 X 500 | 300 | 300 | 27 bytes. | - |
| **ARA** | 50 | 500 X 500 | 300 | 300 | 1 packet of 64 KB /s | - |
| **PACONET** | 100 | 500 X 500 | 120 | 300 | 4 packets of 64 KB /s | - |
| **Ant-AODV** | 100 | 500 X 500 | 600 | 300 | 6 packets of 1000 bytes /s | - |
| **Ant-DSR** | 100 | 500 X 500 | 600 | 300 | 6 packets of 1000 bytes /s | - |

## 4.2 Discussion

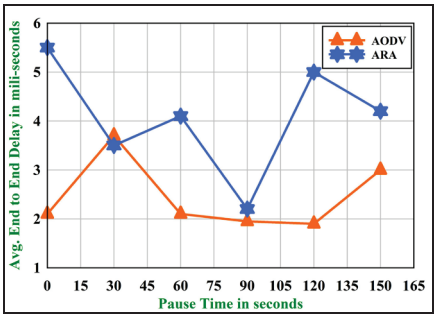
Every trace file generated by NS2 during the simulation of AntNet is of very large size (around 300 MB). Hence, due to this practical difficulty, we have fixed packet size of 27 bytes, so that the trace file size will not increase too much. From Figure 5(a), we can see that the average end-to-end delay for AntNet is better (lower) than AODV, for varying node velocities. From Figure 6(a) we can see that the packet delivery ratio for AntNet is almost equal to 1, whereas for AODV, the packet delivery ratio is less than 1, for varying node velocities. Hence, AntNet performs better than AODV in terms of end-to-end delay and packet delivery ratio.



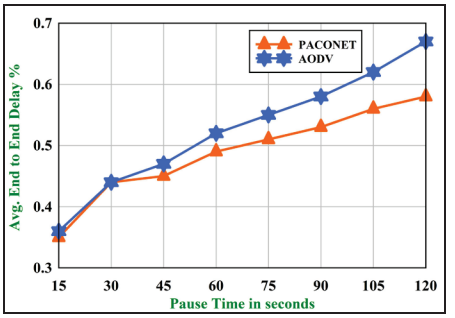

(a)



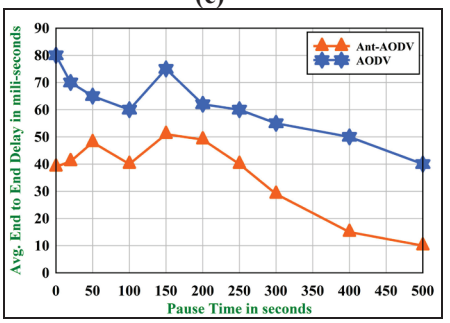

(b)



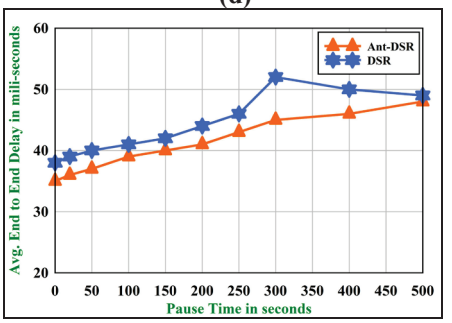

(c)




(d)

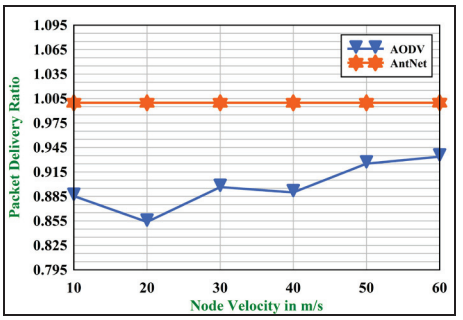



(e)

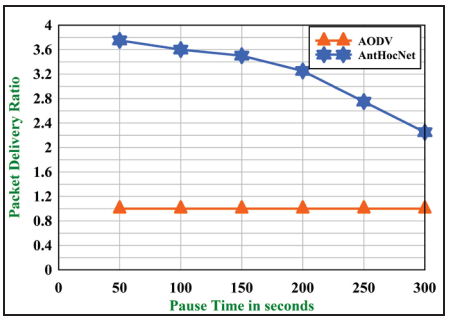



(f)

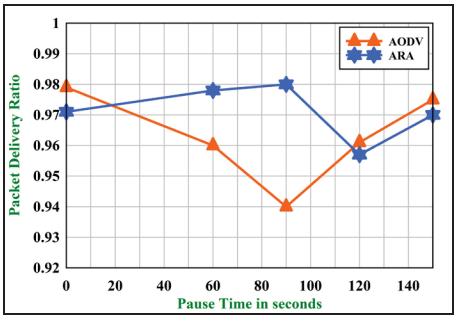
**Figure 5: Average end-to-end delay for: (a) AntNet (b) AntHocNet (c) ARA (d) PACONET (e) Ant-AODV (f) Ant-DSR.**



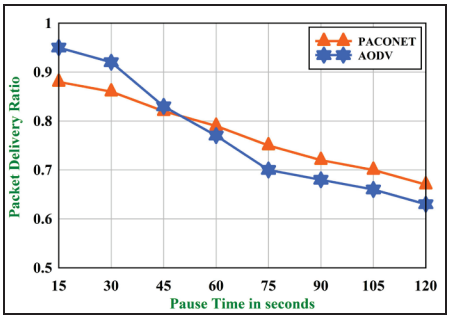

(a)



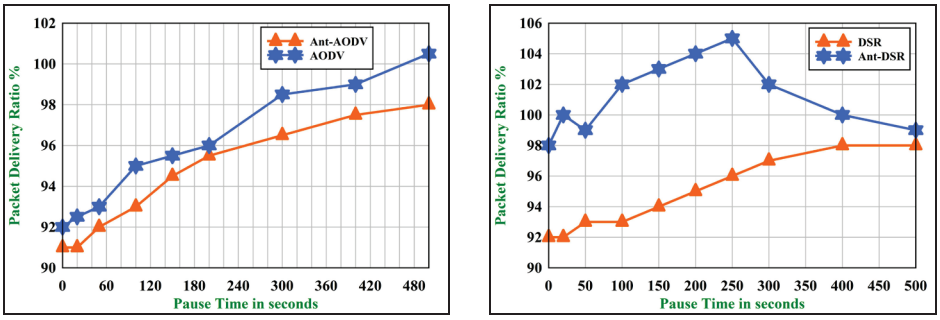

(b)




(c)




(d)




(e) (f)

**Figure 6: Packet delivery ratio for: (a) AntNet (b) AntHocNet (c) ARA (d) PACONET (e) Ant-AODV (f) Ant-DSR.**

From Figure 5(b), we can see that AntHocNet has lower average end-to-end delay as compared to AODV. It is observed that for higher pause times, the end-to-end delay increases. This is due to the fact that under the Random Waypoint model, the nodes are concentrated more in the centre of the network rather than at the edges, especially for lower pause times. From Fgure 6(b), it can be seen that AntHocNet outperforms AODV in terms of packet delivery ratio. This is because AntHocNet uses a lot of different kinds of ant packets, such as FANT, BANT, etc., in order to adapt to changing MANET environments and form many optimal routes, thereby reducing the number of packet drops.
From Figure 5(c), we can see that the end-to-end to end delay for AODV is smaller than ARA. This is due to the fact that while AODV uses no route maintenance mechanism other than a timeout to delete stale routes, ARA uses a route maintenance mechanism to gradually modify the "freshness" of the routes. In addition, the route selection exponent makes the ant route selection equations more sensitive to changes in pheromone values. This change in pheromone values is indirectly indicative of the topology of the MANET and causes the ant route selection equations to select varied routes. Hence, from Figure 6(c), we can see that the packet delivery ratio of ARA is better than that of AODV. For lower mobility, AODV performs better than ARA.

In Figure 5(d), where the node pause time is varied, AODV shows higher overall delay. The delay in PACONET is usually significant at the start of the simulation because of the initial search for routes. Intermediate nodes in AODV are able to respond to route requests, thus saving time for path discovery. In PACONET, the source node has to wait until it gets a BANT sent by the destination. The initial peak in end-to-end delay is what causes AODV to outperform PACONET at higher node speeds. Furthermore, unlike most protocols, PACONET does no broadcasting which limits the extent of path discovery. In Figure 6(d), where the packet delivery ratio is measured at varying node pause times, both algorithms show improved performance. Their performance actually alternates; AODV

starts out better but PACONET finishes with higher delivery ratio for longer pause times.

It is evident from the simulation results shown in Figures 5(e) and 6(e) that by combining ant-like mobile agents with the on-demand route discovery mechanism of AODV, the Ant-AODV hybrid routing protocol would give reduced end-to-end delay and high packet delivery ratio at large pause times. The high packet delivery fraction in Ant-AODV is because it makes use of link failure detection and route error messages. Whereas in the case of ant-based routing, there is no such feature and hence, the source nodes keeps on sending packets unaware of the link failures. This leads to a large amount of data packets being dropped, which reduces the packet delivery fraction and the throughput. It can be observed that end-to-end delay is considerably reduced in Ant-AODV as compared to AODV. As ants help in maintaining high connectivity in Ant-AODV, the packets need not wait in the send buffer until the routes are discovered. Even if the source node does not have a ready route to the destination, due to the increased connectivity at all the nodes, the probability of it receiving replies quickly from nearby nodes is high, resulting in reduced route discovery latency. Ant-AODV is able to provide reduced end-to-end delay and high connectivity as compared to AODV. As a result of increased connectivity, the number of route discoveries is reduced and also the route discovery latency. This makes Ant-AODV suitable for real-time data and multimedia communications. Table 3 summarises the differences between AODV and Ant-AODV.

**Table 3: Comparison between AODV and Ant-AODV.**

| Parameters | AODV | Ant-AODV |
|---|---|---|
| Routing type | Purely reactive | Hybrid |
| End-to-end delay | High | Low |
| Connectivity | Low | High |
| Route type | Single path | Multipath |
| Overhead | Low | High |

From Figure 5(f), we can see that the average end-to-end delay is reduced with Ant-DSR, while from Figure 5(f), we can see that the packet delivery ratio for Ant-DSR is higher than that of DSR. This is mainly due to the addition of the delay pheromone in the RREQ and RREP packets. The reduction in delay is at its maximum (15-20 %) when the pause time reaches beyond 300 s. Both protocols have the same delay for higher pause times. The packet delivery ratio shows an improvement over DSR, where it is high for low pause time. It can be seen that an increase in node speed results in significant decrease in both the protocols due to more link breakages. We can say that Ant-DSR produced better results than DSR in terms of packet delivery ratio and end-to-end delay. Table 4 summarises the differences between AODV and Ant-AODV.

**Table 4: Comparison between DSR and Ant-DSR.**

| Parameters | DSR | Ant-DSR |
|---|---|---|
| Routing type | Reactive | Reactive |
| End-to-end delay | High | Low |
| Energy and jitter | Low | High |
| Through | Low | High |
| Overhead | Low | High |

### 4.3 Summary of ACO Based Protocols for MANETs

Table 5 summarises the ACO based protocols used in this study in terms of routing and path types. Proactive routing protocols maintain routes to all destinations, regardless of whether or not these routes are needed. In order to maintain correct route information, a node must periodically send control messages. Therefore, proactive routing protocols may waste bandwidth. The main advantage of this category of protocols is that hosts can quickly obtain route information and establish a session. Reactive routing protocols can dramatically reduce routing overhead because they do not need to search for and maintain the routes on which there is no data traffic (Singla & Kakkar, 2010).

**Table 5: Summary of ACO based protocols used in this study.**

| Protocol | Routing type | Path type |
|---|---|---|
| AntNet | Proactive | Single |
| AntHocNet | Hybrid | Single |
| ARA | Reactive | Multipath |
| PACONET | Reactive | Single |
| Ant-AODV | Hybrid | Multipath |
| Ant-DSR | Reactive | Broadcast |

Hybrid methods combine proactive and reactive methods to find efficient routes, without much control overhead. In general, hybrid routing's flexibility allows the network operator to adjust the protocol operation to match the network's current mission and state. For example, a purely proactive operation might be used in relatively static networks, such as inter-ship links. In contrast, purely reactive routing might be used in dynamic networks such as clouds of tactical unmanned aerial vehicles (UAVs), and networks of ground-based sensors that have strict low probability of detection (LPD) requirements. These protocol adjustments could occur without changing the network software or "rebooting" any of the underlying MANET routers (Sholander *et al.*, 2002).

Single path routing is based on single route establishment between sourceand destination. In this routing, the packet is transmitted to the destination using a single route (Abolhasan *et al*.,, 2004). Multipath routing gives the choice to the source to choose the path between various available paths between source and destination by taking advantage of the connectivity redundancy of the underlined network (Di Caro, 1998). Broadcast routing is when a single device is transmitting a message to all other devices in a given address range. This broadcast could reach all hosts on the subnet, all subnets, or all hosts on all subnets. Broadcast packets have the host (and/or subnet) portion of the address set to all ones (Abolhasan & Dutkiewicza, 2004).

## 5. CONCLUSION AND FUTURE WORK

In this paper, we reviewed a few ACO based protocols, namely Antnet, AntHocNet, ARA, PACONET, Ant-AODV and Ant-DSR. We obtained QoS parameters such as end-to-end delay and packet delivery ratio for these protocols. We also compared these protocols with conventional protocols, AODV and DSR, based on the random waypoint mobility model. Our results shows that ACO based protocols perform better than conventional protocols in terms of end-to-end delay and packet delivery ratio.

Our research findings may be useful for researchers who wish to modify the existing ACO based protocols. The results obtained in this research can be used for comparison with the modified protocols.
For future work, the performance of these protocols can be analysed using other mobility models such as pursue, random direction and nomadic community mobility models. More critical performance evaluations of these protocols shall be done on the basis of simulations and other performance metrics such as routing overhead, route cost and normalised routing load.

## ACKNOWLEDGEMENT

The authors are grateful to the management of Fr. Conceicao Rodrigues College of Engineering, Fr. Agnel Technical Education Complex, Bandra (W), Mumbai: 400050, for allowing to use their infrastructure for experimentations.

## REFERENCES

Abdel-Moniem, A.M. & Hedar, A. (2010). An ant colony optimisation algorithm for the mobile ad hoc network routing problem based on AODV protocol . *Tenth Intl. Conf. Int. Syst. Design App.*, pp. 1332-1337.

Gupta, A., Sadawarti, H. & Verna, A. (2012). MANET routing protocols based on ant colony optimisation. *Intl. J. Modeling & Optimisation.*,2(1): pp. 42-49.

Abolhasan, M., Wysocki, T. & Dutkiewicz, E. (2004). A review of routing protocols for mobile ad hoc networks. *J. Ad Hoc Networks*, **2**: 1-22.

Aissani, M,. Fenouche, M. & Sadour , F. (2007). Ant-DSR: Cache maintenance based routing protocol for mobile ad-hoc networks. *Third Adv. Int. Conf. Telecom.,* pp. 35-40.

Baluja, S., & Davies, S. (1998). Fast probabilistic modeling for combinatorial optimisation. *Fifteenth National Conf. AI,* pp. 469-476.

Camp, T., Boleng, J. & Davies, V. (2002). A survey of mobility models for ad hoc network research. *Wirel. Commun. Mobile Comput.*, **2**: 483-502.

Caro, G.D. & Dorigo, M. (1998). Mobile agents for adaptive routing. *Thirty-First Hawaii Intl. Conf. Syst. Sci.*, pp. 74-83.

Chengyong Liu, L.L. & Xiang, Y. (2008). Research of multi-path routing protocol based on parallel ant colony algorithm optimisation in mobile ad hoc networks. *Fifth IEEE Intl. Conf. Inf. Tech.*, pp. 1006-1010.

Di Caro, G.A. (1998). Two ant colony algorithms for best-effort routing in datagram networks. *Tenth Intl. Conf. Parallel Distrib. Comput. Syts.*, pp. 28-31.

Di Caro, G.A., Ducatelle, F. & Gambardella, L.M. (2005). AntHocNet: An adaptive nature-inspired algorithm for routing in mobile ad hoc networks. *Eur. T. Telecommun.* , **16**: 443–455.

Gunes, M. & Sorges, U. (2002). ARA-the ant colony based routing algorithm for MANETs. *Intl. Conf. Parallel Proc. Workshops*, pp. 79-85.

Gupta, A.K., & Sadawarty, H. (2010). Performance analysis of AODV, DSR and TORA routing protocols. *J. Eng. Tech.,* **2**: 226-221.

Jha, K. Khetarpal, M. & Sharma, M. (2011). A survey of nature inspired routing algorithms for MANETs. *Third IEEE Intl. Conf. Electron. Comp. Tech.*, pp. 16-24.

Lin, N., & Shao, Z. (2010). Improved Ant Colony Algorithm for Multipath Routing Algorithm Research. *Intl. Symposium Intel. Inf. Process. Trusted Comput.*, pp. 651-655.

Mann, F. & Mazhar, M. (2011). MANET routing protocols vs mobility models: A performance evaluation.. *Third Intl. Conf. Ubiq. Fut. Networks*, pp. 179-184.

Middendorf, M., Reischle, F. & Schmeck, H. (2002). Multi colony ant algorithms. *J. Heuristics,* **8***:* 305-320.

Marwaha, S., Tham, C.K. & Srinivasan, D. (2002). A novel routing protocol using mobile agents and reactive route discovery for ad hoc wireless networks. *10th IEEE Intl. Conf. Networks*, pp. 311-316.

Marwaha, S., Tham, C.K. & Srinivasan, D. (2002). Mobile agents based routing protocol for mobile ad hoc networks. *Global Telecomm. Conf.,* pp. 163-167.

Osagie, E., & Thulasiram, R. (2008). PACONET: imProved ant colony optimisation routing algorithm for mobile ad hoc NETworks. *22nd Intl. Conf. Adv. Inf. Networking App.*, pp. 204-211.

Perkins, C., Belding-Royer, E. & Das, S. (2003). Ad hoc on-demand distance vector (AODV) routing. RFC 3561. Available online at: http://www.ietf.org/rfc/rfc3561.txt (Last accessed date: 20 September 2012).

Singla, V. & Kakkar, P. (2010). Traffic Ppattern based performance comparison of reactive and proactive protocols of mobile ad-hoc networks, *J. Comp. App. ,* **5**:16-20.

Sujatha, B.R.& Harigovindan, V.P. (2010). The research and improvement of AntNet algorithm. *Second Intl. Asia Conf. Informatics Control, Autom. Robotics*, pp. 505-508.

Sholander, P., Yankopolus A., Coccoli, P. & Tabrizi, S.S. (2002). Experimental comparison of hybrid and proactive routing protocols for MANETs. *MILCOM 2002*, pp. 513-518.

Villalba, L.J.G., Matesanz, J.G., Orozco, A.L.S. & Díaz, J.D.M. (2011). Auto-configuration protocols in mobile ad hoc networks. *Sensors*, **11**: 3562-3666.

Villalba, L.J.G. & Orozco, A.L. (2010). Bio-inspired routing protocol for mobile ad hoc networks. *IET Comm.,* **4**: 2187-2195.

# STUDY ON FORMATION OF MAIN RECIRCULATION REGION OVER A BACKWARD-FACING STEP FLOW AT MODERATE RANGE OF REYNOLDS NUMBERS

Yogeswaran Sinnasamy[1, 3*], Noor Arbiah Yahaya[1], Abdul Aziz Jaafar[2] & Azmin Shakrine Mohd Rafie[3]

[1]Ship Silencing Centre (SSC), Royal Malaysian Navy (RMN), Malaysia
[2]Faculty of Manufacturing Engineering, Universiti Malaysia Pahang (UMP), Malaysia
[3] Department of Aerospace Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), Malaysia

[*]Email: yoges_aero@yahoo.com

## ABSTRACT

*In flow dynamics, the prediction of reattachment length RL of the main recirculation region (MRR) of a backward-facing step (BFS) step flow is essential to find the effect of Reynolds number Re on MRR formation based on streamline-contour diagram (SCD) and their pattern evolution. In this study, a computational fluid dynamics (CFD) software called PHOENICS is used to model BFS flows and perform computations at values of Re between 100 and 3,000. The expansion ratio ER of the domain is 2, and two numerical schemes, which are HYBRID (HY) and UPWIND (UP), are applied. Based on the flow structure, the reattachment points of the MRRs are traced manually. The lengths of the MRRs are computed, and compared with experimental results and computational data from previous studies.*

**Keywords:** *backward-facing step (BFS); main recirculation region (MRR); reattachment length RL; Reynolds number Re; streamline-contour diagram (SCD).*

## 1. INTRODUCTION

The flow of air in gas turbine engine components can be studied using a backward-facing step (BFS) configuration to represent the scenario that could take place when flow from higher velocity or smaller channel enters an area with lower velocity or larger channel. BFS is one of the simplest flow configurations, which has higher tendency to create flow with separation and reattachment flow regions. In aerospace engineering, this kind of flow configuration can be seen in the formation of cooling layers on the surface of gas turbine components, in

which the layer is needed to cool down the components and prevent material melt down due to high temperature operation of the engine (Schäfer *et al*., 2009; Lanzerstorfer & Kuhlmann, 2012).

There have been various numerical studies conducted on BFS flow (Denham & Patrick, 1974; Armaly *et al.*, 1983; Lee & Mateescu, 1998; Erturk, 2008; Rouizi *et al*., 2009; Yogeswaran, 2011), whereby varying values of expansion ratio *ER*, which is the ratio between downstream and upstream channel heights, and dimensions of BFS domains have been used. For example, Armaly *et al.* (1983) used a BFS domain with upstream and downstream channel heights of 0.52 and 1.01 cm respectively, producing *ER* =1.94. Erturk (2008) used a computational BFS domain with upstream and downstream channel heights of *h* and 2*h*, producing *ER*=2, where *h* is the step height and is a dimensionless variable. He then decreased the height of the downstream channel to 1.94*h*, producing a second BFS with *ER* = 1.94. At *ER* = 2, he performed computations at *Re* between 100 and 3,000, while at *ER* = 1.94, computations were performed at *Re* between 100 and 1,500. Based on our observation of the related literatures, we found that even though varying dimensions of experimental models and computational domains were used by various researchers, the value of *ER* is almost same and approximately between 1.94 and 2.

Among the techniques employed by these researchers are laser-Doppler anemometer (LDA) measurement (Armaly *et al*., 1983), multi-element hot-film sensor array (MHFS) (Lee & Mateescu, 1998) and numerical computational with a self-developed computational fluid dynamics (CFD) code (Erturk, 2008). The data from Denham & Patrick (1974) was found to be one of the oldest data among them showing obvious difference. This difference could be because of certain factors such as different computer code used, range of Reynold number *Re* used for the computation, grid meshing setting, number of iterations and under-relaxation factors. We believe that some of these factors contribute directly or indirectly to the some of the disagreements among these researchers. Besides that, the difference among these data could be due to accuracy and error percentage of variables involved in the simulation.

In this study, a CFD software called PHOENICS is used to perform simulations on a BFS model with *ER*=2.0. The simulations are performed at a moderate range of *Re*, which is between 100 and 3,000. Based on the flow structure, the reattachment point of the main recirculation streamline of each simulation is identified. The results obtained are then compared with experimental and computational data available in the literature.

## 2. METHODOLOGY

### 2.1 BFS Model

In this study, we designed a BFS model based on the work performed by Erturk (2008), as it meets our basic requirement, our model $ER$ = 2.0 and with larger upstream and downstream channels. Figure 1 shows the schematic of the BFS flow model used in this study. The model has an overall dimension of 320$h$ (length) x 2$h$ (height), where $h$ is the step height. The discontinued lines indicate that the lengths of downstream $L_e$ and upstream $L_x$ channels are 20 and 320 times respectively larger than $h$. Based on 2D laminar flow assumptions, fluid flows with no spanwise velocity components and any variations of kinematic properties in that direction are neglected.

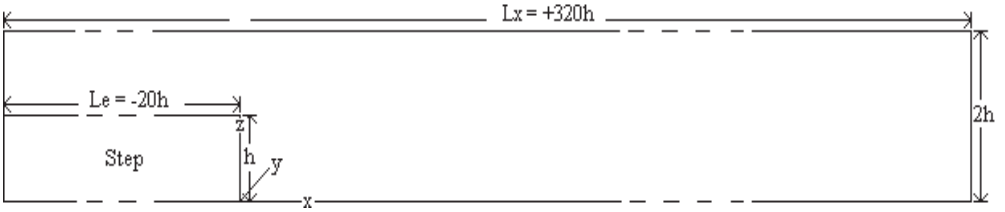



**Figure 1: Geometrical properties of the BFS flow model used in this study.**

### 2.2 Computational Blocks and Grids

The building of computational blocks take into consideration certain factors such as the predicted areas inside the flow channel which could be the spots where main recirculation region (MRR), and subsequent regions or bubbles are present when the air from the inlet flows along the upstream channel and enters the downstream channel after passing the edge of the step. Based on Armaly *et al.* (1983) and Erturk (2008), Figure 2 is sketched to show the predicted areas of recirculation regions were expected to be formed. Based on this figure, we can determine approximately the locations and sizes of the recirculation regions when it passes the edge of the step and heads to the outlet of the channel. There are actually more recirculation regions that could be found beyond the ones shown in Figure 4. However, this was not considered in this study because it basically concentrates on the MRR and the reattachment length. We will only discuss briefly on the effect of first and second subsequent recirculation regions on the MRR.

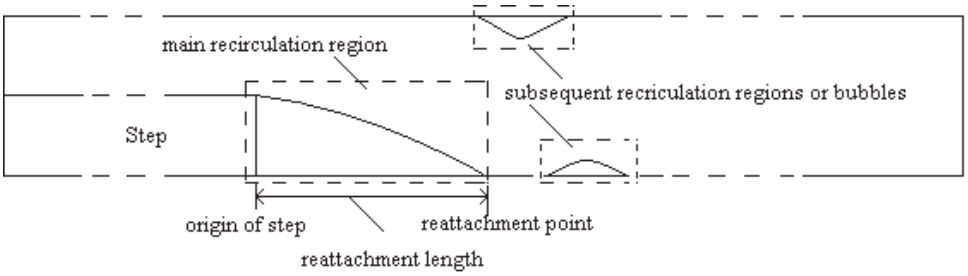



**Figure 2: Predicted spots for the main recirculation region and subsequent recirculation regions for the BFS model.**

Based on Figure 3, this study only focuses on 2D flow, whereby the domain is divided into nine regions; in which four regions are in the *x*-direction or horizontal plane, and five regions are in the *z*-direction or vertical plane. We added an additional region as shown in Table 1 which represents the *y*-plane in our BFS domain. This region represents the thickness or width of the domain, which is 0.01 times smaller than *h*.

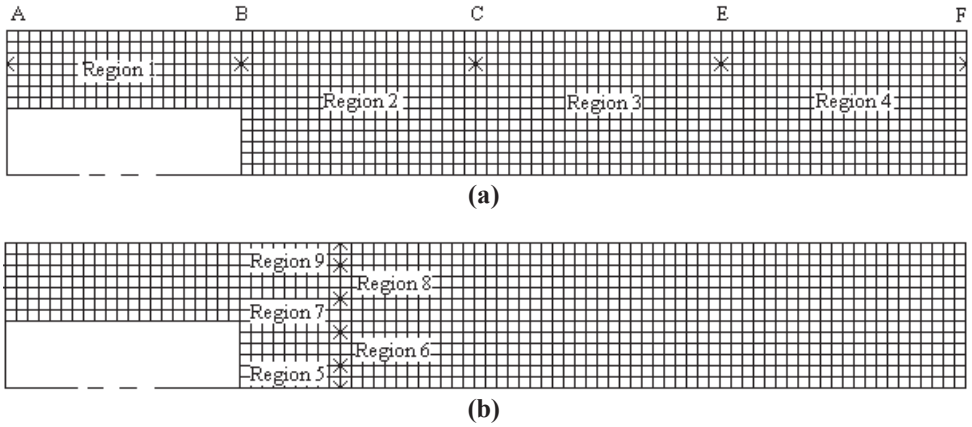



**Figure 3: An example of computational grids arrangement for the BFS domain; (a) *x*-direction and (b) *z*-direction.**

Besides the regions and their identification number, Figure 5 also shows the domain after the meshing or gridding process in the *x*- and *z*-planes. This is not the actual number of grids in our study and the purpose of showing this figure is to assist in our explanation later. The details of computational grid properties for the BFS flow domain are tabulated in Table 1. The table shows that the number of grids is meshed up to 2425 (*x*) x 100 (*z*) x 1(*y*). Therefore, the total number of grids is 242,500. This large number of grids ensures that our computational domain has a certain level of accuracy in terms of identifying the reattachment point of the MRR.

**Table 1: Computational grid properties for the BFS flow model.**

| Region | End Point | Cells | Distribution | Power | Symmetric | Cell Power |
|---|---|---|---|---|---|---|
| ***x*-direction** | | | | | | |
| 1 | 0.188 | 250 | Geometric Program | 1.000 | No | Set |
| 2 | 1.130 | 1250 | Geometric Program | 1.000 | No | Set |
| 3 | 2.072 | 300 | Geometric Program | 1.020 | Yes | Set |
| 4 | 3.014 | 625 | Geometric Program | 1.000 | No | Set |
| **Total** | | **2425** | | | | |
| ***z*-direction** | | | | | | |
| 5 | 0.0025 | 20 | Geometric Program | 1.100 | No | Set |
| 6 | 0.0075 | 20 | Geometric Program | 1.000 | No | Set |
| 7 | 0.0120 | 20 | Geometric Program | 0.950 | Yes | Set |
| 8 | 0.0170 | 20 | Geometric Program | 1.000 | No | Set |
| 9 | 0.0190 | 20 | Geometric Program | -1.100 | No | Set |
| **Total** | | **100** | | | | |
| ***y*-direction** | | | | | | |
| 10 | 0.010 | 1 | Power Law | 1.000 | No | Free |

Geometric contraction and expansion factors are then applied on each region using the values given in Table 1. All the regions in the in *x*-plane are applied with a factor of 1.0. This means that all the grids of the *x*-plane will have no contraction and expansion characteristics, and stay with original size and properties of the grids. Meanwhile, grids in regions of the *z*-plane applied with are factors between -1.1 and 1.1, which is related with the predicted spots of the MRR and subsequent recirculation regions as indicated in Figure 4. For example, the factor value with negative term, such as -1.1, means that the grids in the respective regions have been imposed with contraction effect. On the other hand, the factor value with positive term such as 0.950, 1.0 and 1.1 means that the grids at the respective regions have been imposed with expansion effect.

## 2.3 Boundary Conditions of the BFS Model

Figure 4 shows the implementation of boundary conditions for the BFS model. The inlet boundary is located at $L_e$ = -20$h$, while the outlet boundary is located at $L_x$ = +320$h$. The fluid enters the main channel with no traverse velocity with uniformly distributed streamwise velocity component. At all wall surfaces, no-slip boundary conditions are applied. Pressure boundary conditions are applied at the outlet boundary to avoid any influx of air entering the BFS model.

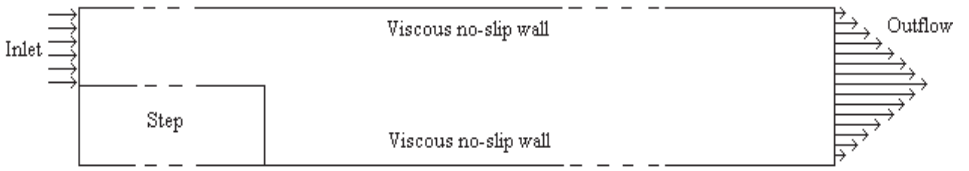



**Figure 4: Fluid flow and boundary conditions applied for the BFS model.**

## 2.4 Convergence Criteria and Numerical Schemes

In this study, the effect of different numbers of iterations on the flow variables will not be covered. This study on considers the effect of different values of *Re* on the formation of flow streamlines and their reattachment length *RL* of the MRR. Therefore, in this study, we will use the maximum number of iterations that is considered sufficient for the flow to converge which is 3,000. The convergence criterion is set at 1.0 x $10^{-6}$ to ensure accuracy in our flow solution. In addition, accuracy of the solution also depends on the numerical schemes used in the study. Numerical schemes play an important part in aspects that involves the numerical stability and boundedness of the solution (Malin & Waterson, 1995).

In this study, the HYBRID (HY) scheme (also known as the hybrid-differencing scheme) is used. HY is a default scheme which employs the first-order UPWIND (UP) scheme (also known as the upwind differencing scheme) in high convection regions and the second-order central-differencing scheme (CDS) in low convection regions. UP is bounded and highly stable, but is highly diffusive when the flow direction is skewed relative to the grid lines. HY is only marginally more accurate than UP, as the second-order CDS will be restricted to regions of low Péclet number, which is the ratio of the rate of advection of a physical quantity by the flow to the rate of diffusion of the same quantity driven by an appropriate gradient (Malin & Waterson, 1995).

Once all the computation procedures and details are entered into the PHOENCIS software, the next step will be the execution of the computation, and the flow will be monitored. One of the examples of computation window which shows the

flow details and residuals of variables involved is shown in Figure 5. The flow is considered as converged when the residual plots, which are indicated with three different colours on the right hand side of the computation window, shows no more signs of instability.

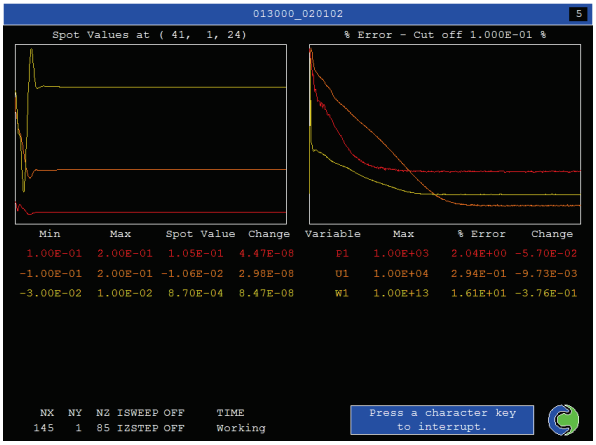



**Figure 5: An example of one of the computation window showing the flow details and residuals of variables.**

## 3. RESULTS AND DISCUSSION

For this study, we obtained 60 streamline-contour diagrams (SCDs), which represent 30 cases of HY and UP schemes respectively. Each scheme consists of SCDs with values of *Re* from 100 to 3,000. Based on these SCDs, *RL* values are extracted manually from the MRR of each diagram and tabulated to establish a database. This database will be used for comparison with results from previous studies.

### 3.1 SCD Interpretation

For each SCD, the number of streamlines needs to be decided appropriately because it is an important aspect in setting the resolution of the images formed and will eventually influence the interpretation. For this study, the number of streamlines used is 10, which gives better visual appearance in comparison to when fewer numbers of streamlines are applied. While the number of streamlines can be increased, it would increase the computation time for the SCDs.

Examples of the generated SCDs are shown in Figure 6. Each SCD shows the flow of air represented by streamlines. The streamlines inside the upstream

channel are straight lines, with no interaction between lines. This shows that the streamlines are constant with no turbulence behaviour in the flow. The SCDs are used in identifying recirculation regions on the bottom and upper surfaces as shown in Figure 2.

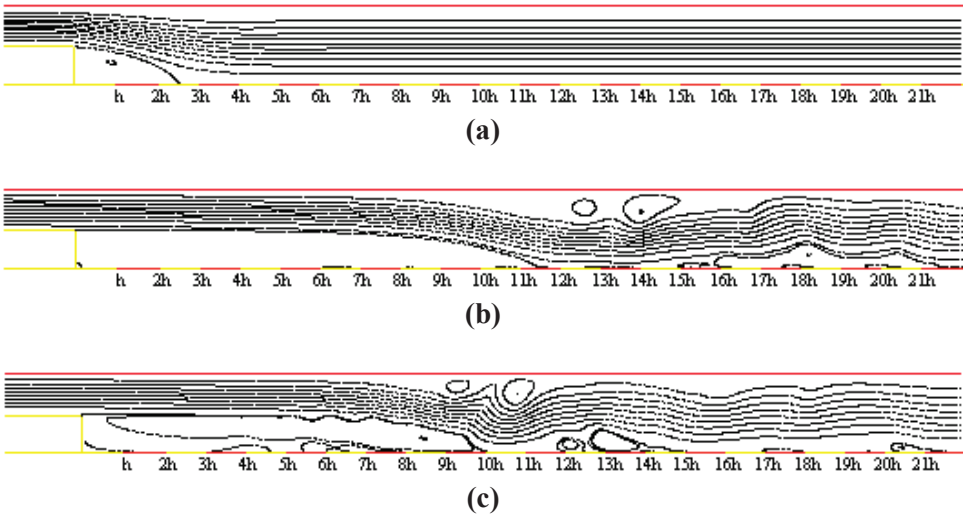



**Figure 6: Streamline contours for *Re* = (a) 100, (b) 1,500 and (c) 3,000.**

The value of *RL* is obtained based on the distance measured between the origin of the step and point of reattachment from the inlet of the channel and flow along the upstream channel after passing by the edge of step. When we make comparison between these two channels, we can see a lot of difference in terms of distribution of the streamlines whereby the gap between them becomes slightly larger and the formation of recirculation regions that happens at both bottom and upper surfaces. At lower range of *Re,* for example, between 100 and 400, once the flow passes the edge of the step, the streamlines which are nearest to the bottom surface of the upstream channel are only able to form the MRR which is located behind the step. This region is formed in such a way that the streamlines passing the edge of the step settle down on the bottom surface of the downstream channel and formed a curve or arch-like smooth line. One of the examples is shown in Figure 6 (a). However, when *Re* is increased to above 400, we found more recirculation regions formed on both surfaces with the MRR having increasing length.

### 3.2 *RL* Values Obtained for Present Study

The values of *RL* of our study are shown in Figure 7. We found that the trendings of RLs for both the HY and UP schemes are almost similar. When we use the HY scheme, we found that RL increases in a proportional manner when the *Re* is increased from 100 to 1,200. At *Re* = 1,300 and above, we found RL decreases with fluctuating trend.

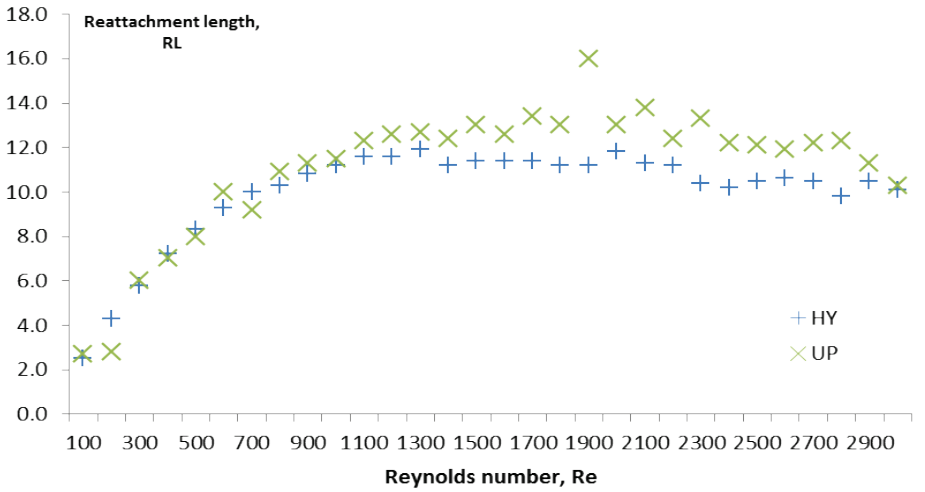



**Figure 7: Comparison of RL values of present study for HY and UP schemes.**

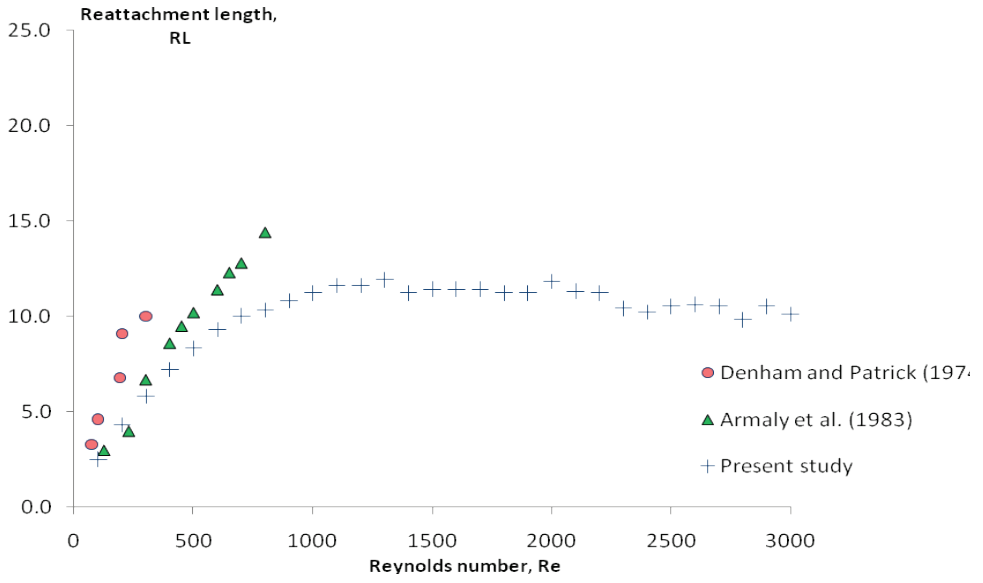
When we used the UP scheme, we found the same trending as HY, where at the lower range of *Re*, *RL* increases when the *Re* is increased as well. However, when *Re* is increased from 700 to 1,200, we found that the rate of increment of *RL* starts to decrease. The rest of trending of *RL* which is between 1,300 and 3,000 seems to be unpredictable because *RL* always oscillates between 10.3 and 16.0. For example, at *Re* = 1,300, 1,900, 2,600 and 3,000, RL is 12.7, 16.0, 11.9 and 10.3 respectively. This shows that the *RL* at the moderate range of *Re* is inconsistent due to decreasing size of the MRR.

We conclude that based on these findings, the formation of the bottom and upper recirculation regions has impact on the size of the MRR when *Re* is increased from 100 to 3,000. At lower range of *Re*, there is no large impact, but it is obvious when for values of *Re* between 1,300 and 3,000, which is considered as moderate range of *Re*. While at lower range of *Re*, *RL* increases linearly, the trending starts to show some decrement as *Re* is further increased.

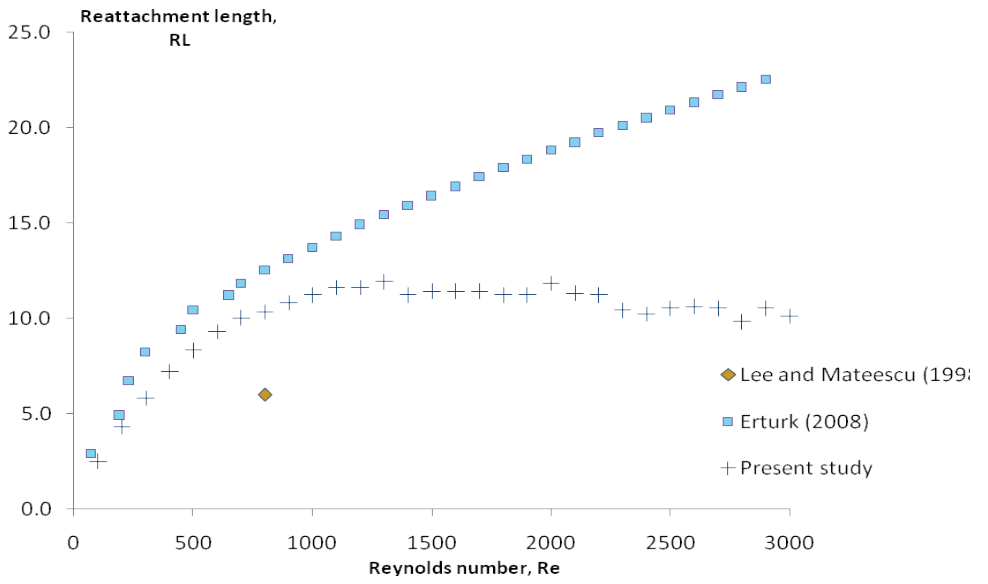
We found that *RL* is decreases when *Re* to more than 3,000. It seems to be that the growing of MRR becomes is destabilised by the growing of the subsequent recirculation regions on the bottom and upper surfaces inside the channel.

### 3.3 Comparison of Present Study With Previous Studies

Figure 8 shows the comparison in plots of *RL* vs *Re* for the present study with previous studies. Figures 8(a) and (b) show comparison with data obtained via experimental and computational methods respectively.




**(a)**




**(b)**

**Figure 8: Comparison of plots of *RL* vs *Re* between our study and (a) Denham & Patrick (1974) and Armaly *et al.* (1983) (b) Lee & Mateescu (1974) and Erturk (2008)**

### 3.3.1 Denham & Patrick (1974)

Denham & Patrick performed their experimental work at *Re* = 73, 100, 191, 200 and 300. Based on our analysis, we found that both trendings of *RL* values are increasing when the *Re* is increased as well. For example, at *Re* = 100, they obtained 4.6, while we obtained 2.5. At *Re* = 200, they obtained 9.1, while we obtained 4.3. Based on these initial stage of data observation, we found the existence of effect of *Re* on size and growing of MRR. At *Re* = 300, they obtained *RL* equal to 10, while we obtained 5.8. We found that *RL* obtained by Denham & Patrick is 1.72 times larger than ours at this final *Re*. We also found that the *RL* values obtained by them are two times larger than ours. This can be proven based on the trending line gradient of RL values of Denham & Patrick which is higher than ours for the mentioned *Re* values as above. We conclude that both trendings exhibit only small scale of similarities especially at the low range our data of *RL*.

### 3.3.1 Armaly *et al.* (1983)

Armaly *et al*. performed their experimental work at *Re* = 125, 229, 300, 400, 450, 500, 600, 650 700 and 800. Based on our analysis, we found differences in terms

of *RL* values exist between these data. We also found that *RL* values obtained in Armaly *et al*. increased nonlinearly, while *RL* values obtained in our study is increased in proportional manner when the *Re* is increased.

We compared that results at *Re* = 300 and found the *RL* value obtained by Armaly *et al*. is 6.7, while we obtained 5.8. Even though the *RL* value obtained by Armaly *et al*. is 1.16 times larger than ours, we considered it as satisfactory. At *Re* = 400 and 500, they obtained 8.6 and 10.2 respectively, while we obtained 7.2 and 8.3 respectively. We found that at all three *Re*, the results obtained by Armaly *et al*. are larger than our results, but within the same range of difference. However, these results show some good agreement when compared with Denham & Patrick (1974).

At *Re* = 600, Armaly *et al*. obtained 11.4, while we obtained 9.3. Finally at *Re* = 700 and 800, they obtained 12.8 and 14.4 respectively, and we obtained 10.0 and 10.3 respectively. From these results, we found that *RL* values obtained by Armaly *et al*. keeps on increasing with larger scale, while our results shows a sign of slowing down in the when the *Re* is increased. Based on these findings, we conclude that the results of our study are more agreeable with Armaly *et al*. in comparison with Denham & Patrick (1974).

### 3.3.2 Lee & Mateescu (1998)

Lee & Mateescu performed computation at *Re* = 800 and obtained *RL*=6.0, while we obtained *RL*=10.3. Even though only one case of computation was performed by them, we can still see some relationship between the two studies. We found that the *RL* value obtained by Lee & Mateescu is 1.72 times smaller than ours. This could be the effect of number of grids, in which our computational domain is finer than Lee & Mateescu (1998). In our study, we used computational domain with number of grids of 242,500, while they used a domain with number of grids of 25,000. Our domain is 9.7 times finer than Lee & Mateescu. Furthermore, they used convergence criteria less than $5 \times 10^{-5}$, while we used convergence criteria $1 \times 10^{-6}$. Therefore, we conclude that the number of grids could be among the factors that affect the computed values of *RL*, and eventually influence the size and growing momentum of the MRR.

### 3.3.3 Erturk (2008)

There are various similarities between our study and Erturk; computations performed for *Re* values between 100 and 3,000 with increment steps of 100, *ER*=2, and similar lengths of upstream and downstream channels. There are also some differences between both studies, such as in terms of number of grids, in which number of grids used by Erturk is 429,351 and in our study, it is 242,500.

This shows that the computational domain used by Erturk is finer than the domain used in our study.

We found that at lower range of *Re*, both studies show similar trendings in terms of increase of *RL*. For example, at *Re* = 100, RL values obtained Erturk and us is 2.9 and 2.5 respectively. This shows a close and well agreed RL between both of us. At *Re* = 200, he obtained 4.9, while we 4.3. Once again, we can see a close relationship between the RL values. At *Re* = 300, the *RL* values in Erturk's and our work is 6.7 and 5.8 respectively. This is also indicates good agreement between Erturk's and our data.

However, at *Re* = 400, both sets of data start to show a larger difference in *RL*, in which Erturk obtained 8.2, while we obtained 7.2. This gap of difference between both set of data keeps on increasing and becomes more obvious when the *Re* is further increased. For example, at *Re* = 700 and 1,000, he obtained 11.2 and 13.1 respectively, while we obtained 10.0 and 11.2 respectively. This shows a clear and obvious disagreement between both sets of data when *Re* exceeds the lower range.

At *Re* = 1,500, Erturk obtained 15.9, while we obtained 11.4. At this stage, we found that the *RL* values obtained by Erturk continues increasing and does not show any consistency. The increment gradient of RL values obtained in our study shows a slight decrease when *Re* is increased from 1,000 to 1,500. We found that after *Re* =1,300, the trending of RL values in our study is fluctuates at a low amplitude. This is an indication of the trending line heading towards consistency when the *Re* is increased furthermore.

Even though at the early stage or at low range of *Re,* there is a good agreement between Erturk's and our data, we found that further increase in the *Re* makes a lot difference between both results. This could be dues to the difference in terms of number of grids and numerical schemes employed by both studies. The difference between Erturk's and our studies could be minimised if more computations are performed and analysis on other aspects related with computation, such as number of iterations and under relaxation-factors, is conducted.

## 4. CONCLUSION

In this study, we computed RL values of MRRs of a BFS model and their subsequent separation and reattachment regions on bottom and upper surfaces. Based on these results, we managed to plot the trending of *RL* of two numerical schemes, HY and UP, for moderate range of *Re*. These numerical predictions were compared with experimental and computational data from previous studies to evaluate the accuracy of our computational model in terms *RL* values. The HY

and UP schemes available in PHOENICS have been assessed in terms of their convective and diffusion terms on the size and location of recirculation regions. Other high-order schemes, especially QUICK, should be tested in terms of their accuracy in presenting better and more convincing BFS flows.

## REFERENCES


Armaly, B.F., Durst, F., Pereira, J.C.F & Scho¨nung, B. (1983). Experimental and theoretical investigation of backward-facing step flow. *J. Fluid Mech.*, **27**: 473–96.

Denham, M.K. & Patrick, M.A. (1974). Laminar flow over a downstream-facing step in a two-dimensional flow channel. *Trans. Inst. Chem. Engrs.*, **52:** 361**.**

Erturk, E (2008). Numerical solutions of 2-D steady incompressible flow over a backward-facing step, Part I: High Reynolds number solutions. *Comput. Fluids*, **37**: 633–655.

Lanzerstorfer, D. & Kuhlmann, H.C. (2012). Global stability of the two-dimensional flow over a backward-facing step. *J. Fluid Mech.*, **693**: 1-27.

Lee, T. and Mateescu, D. (1998).Experimental and numerical investigation of 2-D backward-facing step flow. *J. Fluids Struct.*, **12**: 703-716.

Malin, M.R. & Waterson, N.P. (1995). *Schemes for Convection Discretisation in PHOENICS*. CHAM Limited, Bakery House, London.

Rouizi, Y., Favennec, Y., Ventura, J. & Petit, D. (2009). Numerical model reduction of 2D steady incomressible laminar flows: Application on the flow over a backward-facing step. *J. Comput. Phys.*, **228**: 2, 239-2,255.

Schäfer, F., Breuer, M. & Durst, F. (2009). The dynamics of the transitional flow over a backward-facing step. *J. Fluid Mech.*, **623**: 85-119.

Yogeswaran, S. (2011). Assessment on effects of under-relaxation factors on 2D incompressible laminar flow over a backward-facing step (BFS). *Defence S&T Tech. Bull.*, **4**: 39-49.

# NUMERICAL SIMULATION OF RUBBER PANEL UNDER IMPACT LOADING

Kamsani Kamal*, Md Zaini Zainal, Mustafa Omar, Osmera Ismail, Raynee Ramliza Raybayi, Mohd Redhuan Mohd Isa, Azmi Minal, Mohd Fauzy Mohd Nor & Nor Afizah Salleh

Weapons Technology Division (BTP), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*E-mail: kamsani.kamal@stride.gov.my

## ABSTRACT

*This paper presents a study using numerical simulation of bullet impact on rubber panels in order to investigate the depth of penetration into the panels. The study was conducted using a finite element analysis (FEA) software, MSC Dytran. The simulation was conducted with rubber panel of varying thicknesses and values of elastic coefficient Constant A (CA) in the Mooney-Rivlin rubber model. The simulation results show that the rubber panel with thickness of 500 mm and CA of 1.4* x *0.89 MPa is suitable to be used as bullet stopper because it allows the bullet to penetrate and be trapped bullet inside it. Based on the findings of this study, it can be concluded that rubber can be considered as a potential material to be selected as bullet stopper provided that the appropriate geometry and configuration of the panel are chosen.*

**Keywords:** *Depth of penetration; rubber panel; bullet stopper; finite element analysis (FEA); elastic coefficient Constant A (CA).*

## 1. INTRODUCTION

The shooting range of the Weapons Technology Division, Science & Technology Research Institute for Defence (STRIDE) presently uses sand as the material to stop bullets. Sand is also the common material used as bullet stoppers in most of shooting ranges in Malaysia. However, using sand has drawbacks such as being heavy and dusty. The dust will scatter and fly around the shooting range area and reduce the visibility of the target. Therefore, there is a proposal of changing sand with the other materials. There are many types of materials that have been studied, such as aluminium plates (Iqbal *et al.,* 2010), ceramic (Lamberts, 2007; Krishnan *et al*., 2010), composite (Sareen & Smith, 1996; Ong *et al*., 2011), plastic (Dean & Read, 2001; Qasim, 2009), sandwich plates (Zeng, *et al*., 2010), steel (Iqbal *et al*., 2010), polymethylmethacrylate (PMMA) plates (Dorogoy *et*

*al*., 2010), concrete (Nyström & Gyalltoft, 2011), and textile (Barauskas *el al*., 2005; Wang *et al*., 2010).
In this study, the material selected to stop and trap small arms bullets is rubber panels. This selection is as rubber is easily available in Malaysia. Indeed, Malaysia supplies about 46% of the world's rubber supply, and cultivates two to four million acres of rubber, or about 65% of total land used for cultivation (Kishore, 2010). The available supply of this raw material locally has made its price relatively cheap. Therefore, the cost of producing rubber panels is relatively low.

The objective of this study is to determine the depth of penetration of bullets into rubber panels using numerical simulation. This study only considers panel thickness and elastic coefficient Constant A (CA) of the rubber panel. The others parameters, such as elastic coefficient Constant B (CB), bullet velocity, angle of impact, and panel width and length, are left constant. CA and CB are characteristics of rubber after vulcanisation in order to determine its durability (Sombatsompop, 1998). This analysis is aimed at determining the best values of CA and panel thickness which can provide the suitable depth of penetration for bullets not to perforate the rubber panel. Panels with properties which do not allow penetration or allow only partial penetration of bullet length are not considered because the bullets may bounce back, and hit the shooters or equipments.

Depth of penetration studies using numerical simulation was previously employed by Borvik *et al*. (2011) for normal and oblique impacts on 20 mm thick AA6082-T4 aluminium plates. The bullets used were 7.62 x 63 mm NATO soft lead core and 7.63 x 63 mm APM2 hard steel core, with impact velocity of 830 m/s. They found that the angles that can give higher depth of penetration for both bullets are 0 to 60°. Gama & Gillespie (2011) studied the impact, damage evaluation and penetration of thick section composites. They developed a systematic model-experiment methodology to simulate and validate finite element models of progressive composite damage from static to impact loading conditions. They also developed a 3D finite element model for ballistic impact.

The numerical simulation in this study is conducted using finite element analysis (FEA). The basic concept of FEA is to simplify complicated objects by developing small pieces of simple blocks. These blocks are interconnected with each other and any displacement is associated with the other blocks. A displacement function is assigned and connected with every block. Using known properties of stress or strain of the object, the behaviour of nodes in terms of properties of every block can be determined (Logan, 2002). FEA studies on materials under impact have been previously conducted, such as for amour personnel and car protection (Lamberts, 2007), high-strength fabrics for personnel protection (Duan *et al*., 2006), aircraft fuel tanks (Sareen & Smith,

1996), fluid-filled containers (Sauer, 2011) and ceramic armour (Sands *et al*., 2009).

## 2. METHODOLOGY

The models of the bullet and rubber panel for use in the numerical simulation are drawn using the SolidWorks 2007 computer aided design (CAD) software. The data of the bullet and rubber properties are obtained from available references. The simulation modelling is carried out in 3D using the MSC Dytran FEA software with the Lagrangian method solver and the Adaptive Contact Master-Slave Surface failure behaviour command. The Lagrangian method solver uses elements of material with constant mass to connect the grid points on the object forming a mesh. As the body deforms, the grid points will move with the body and the elements (mesh) distort. The Lagrangian method solver then calculates the motion of 27 elements of constant mass (MSC, 2006).

The Adaptive Contact Master-Slave Surface failure behaviour command allows a penetrating object to go through a closed surface after elements in its path have failed, without causing holes in its connectivity. Any contact failure behaviour requires master / slave designations for the contact surface. In this simulation, the bullet is designated as the master and the panel material model as the slave. For every time step, MSC Dytran will check the adaptive contact for the grid points of the bullet to see if any have penetrated into the elements of the panel. If penetration is found, a force is applied on the bullet grid in the direction opposite to the penetration. An equal and opposite force is also applied to the grids which are connected to the element being penetrated. What makes the contact adaptive is its ability to allow the projectile to punch through the hole created when the elements in front of it have failed. When an element in the contact region fails, the adaptive contact algorithm stops applying force to it and from it on the penetrating object (MSC, 2006).

### 2.1 Bullet Model

There are wide varieties of bullet diameters or calibres in the market. The diameter of bullet selected for this study is 7.62 mm. It is heavier in mass than 5.56 mm bullets and has higher velocity than 9 mm bullets. Therefore, it carries the highest energy as compared with 5.56 and 9 mm bullets, and therefore, has higher penetration capability (Savvateev *et al*., 2001). The properties of the bullet used for this study are as shown in Table 1.

**Table 1: Properties of the 7.62 mm bullet (Gander & Cutshaw, 2000).**

| Bullet Type | Muzzle Velocity | Muzzle Energy | Bullet Weight |
|---|---|---|---|
| 7.62 x 51 mm Ball M80 | 854 m/s | 3519 J | 9.65 g |

The model of the bullet illustrated in Figure 1 corresponds to the outer shape dimension of a 7.62 x 51 mm NATO Ball M80 bullet (with a soft lead core). The dimension of the ogive nose part is 76.2 mm in radius, the overall length is 28.63 mm, and the diameter is 7.62 mm (Gander & Cutshaw , 2000).

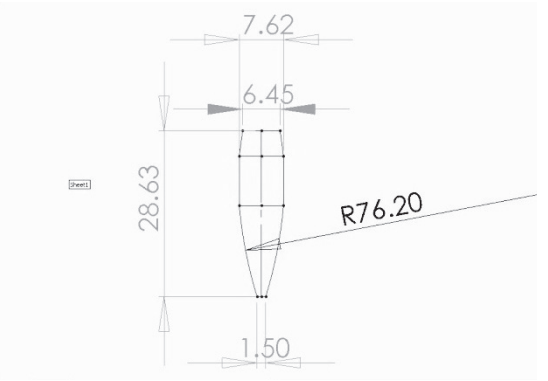



**Figure 1: Dimensions of a 7.62 x 51 mm NATO Ball M80 bullet (Gander & Cutshaw, 2000).**

The mesher of the bullet model input is assigned as the surface type using tetrahedral element shape via the paver meshing algorithm with Tria 3 topology. The total number of shell elements counted is 255. The solid finite element model is shown in Figure 2.

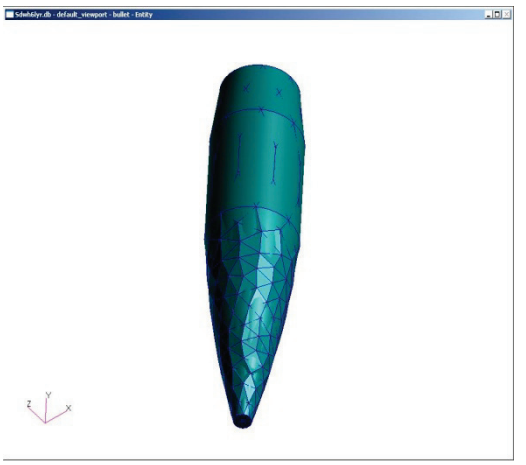


**Figure 2: The solid finite element model.**

The material constitutive model of bullet input command is assigned as rigid (MATRIG) body, which means that the bullet does not deform during impact. The bullet model is assigned as a rigid body to ensure that the panel will receive the highest energy on impact. If the bullet breaks, the energy will reduce, decreasing the depth of penetration. The initial velocity assigned to the bullet model is 800 m/s. This initial velocity is assigned at the nodes of the elements as shown in Figure 3.

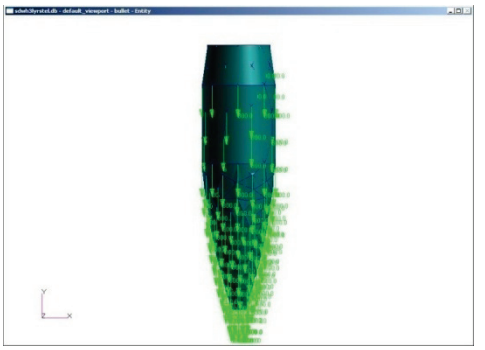


**Figure 3: The initial velocity is assigned at the nodes of the elements.**

## 2.2 Rubber Panel Model

The properties of the rubber material used in this study are as shown in Table 2. The rubber panel model is drawn with dimensions of 100 mm in length and 100 mm in width. Different thicknesses (200, 300, 400 and 500 mm) are used for the model.

**Table 2: Properties of the rubber material used in this study (Roberts & Benzies, 1977).**

| | |
|---|---|
| **Density** | 1.1 kg/m$^3$ |
| **Elastic coefficient CA** | 0.89 Mpa |
| **Elastic coefficient CB** | 0.46 Mpa |
| **Poisson Ratio** | 0.5 or 0.49999 |

The rubber panel is uniformly meshed and assigned as hexahedron eight-noded shape with IsoMesh algorithms input. The mesher is compact at the centre part of panel from the upper to bottom surface. The number of elements used is 8,000. An example of the plot of the mesh of the rubber panel is shown in Figure 4. The boundary condition of the panel is fully

fixed on every finite element node at the edge of the panel. All edges of the rubber are constrained in all degrees of freedom.

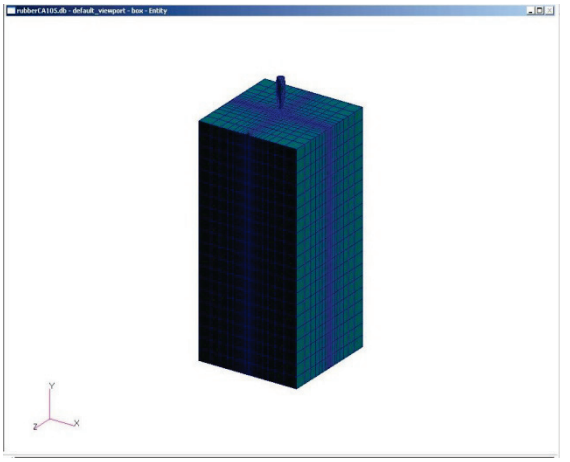


**Figure 4: An example of the plot of the mesh of the rubber panel.**

The material entry of the constitutive model is a user define material model known as rubber (RUBBER 1) with the Lagrangian solid element type. The failure behaviour of a Mooney-Rivlin rubber model is used as it more suitable for hyperelastic material (MSC, 2006). All the settings of the simulation are by default and the rubber panel material is assumed as isotropic.

The strain energy density $\boldsymbol{W}$ function is defined according to the Mooney-Rivlin model (MSC, 2006):

$$\boldsymbol{W}(I_1,I_2,I_3) = \boldsymbol{A}(I_1 - 3) + \boldsymbol{B}(I_2 - 3) + \boldsymbol{C}\left(\frac{1}{I_3^2} - 1\right) + \boldsymbol{D}(I_3 - 1)^2 \quad (1)$$

The elastic coefficient constants $\boldsymbol{A}$ and $\boldsymbol{B}$, and Poissons's ratio $\boldsymbol{v}$ are the input parameters for the model. $I_1$, $I_2$, and $I_3$ are strain invariants in terms of stretches. The constants $\boldsymbol{C}$ and $\boldsymbol{D}$ are related to the input parameters as:

$$\boldsymbol{C} = \frac{1}{2}\boldsymbol{A} + \boldsymbol{B} \quad (2)$$

$$\boldsymbol{D} = \frac{A(5\boldsymbol{v} - 2) + B(11\boldsymbol{v} - 5)}{2(1 - 2\boldsymbol{v})} \quad (3)$$

## 2.3 Depth of Penetration Simulation on the Rubber Panel

In this penetration study, the models are arranged in vertical position. The bullet is positioned above and centre of the panel surface with distance of 2 mm. The angle of impact is set at normal incidence (0° obliquity) to the top surface of the panel. MSC Dytran simulates only a single impact event and is assigned to hit at the centre of the panel in the vertical path as shown in Figure 5. The depth of penetration study is conducted with different values of CA (1.0 x 0.89, 1.2 x 0.89, 1.4 x 0.89, 1.6 x 0.89, 1.8 x 0.89 and 2.0 x 0.89 MPa) for the different thicknesses of the panels (200, 300, 400 and 500 mm).

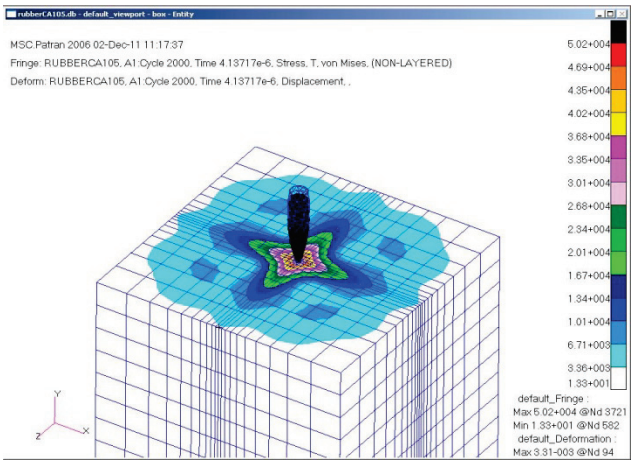



**Figure 5: Simulation of the mechanical response of the rubber material under impact loading condition.**

## 3. RESULTS AND DISCUSSION

The results of the simulations for the rubber panels with different thicknesses are shown in Figure 6. It is observed that for the rubber panels with thicknesses of 200 and 300 mm, the bullet is still able to penetrate and perforate the panels with CA of 1.4 x 0.89 and 1.6 x 0.89 MPa. For the rubber panels with thickness of 300 m, the penetration of the bullet is about 150 mm from the surface of the bullet entry for the panel with CA of 1.8 x 0.89 MPa. For the panels with thickness of 400 mm, the panel with CA of 1.4 x 0.8 9 MPa is penetrated and perforated by the bullet, and for CA of 1.6 x 0.89 MPa, it allows only partial penetration, of about 20 mm, of the bullet into the panel.

For the rubber panels with thickness of 500 mm, the bullet is able to penetrate and perforate the panels with CA of 1.2 x 0.89 MPa and below. For CA of 1.4 x 0.89 MPa, the bullet penetrates the panel and stops at 32 mm from the front of

the panel's surface of the bullet entry. The bullet is unable to penetrate more than 22 mm in depth for the panels with CA of 1.6 x 0.89, 1.8 x 0.89 and 2.0 x 0.89 MPa.

In this simulation, different types of rubber material have been considered. The different types of rubber are known from different values of CA input. The results show that depth of penetration reduces with increasing CA. This is because when the value of CA increases, the stiffness of the rubber panel increases accordingly. Based on the results obtained, it is concluded that the best parameters for rubber panel are CA of 1.4 x 0.89 MPa and thickness of 500 mm.

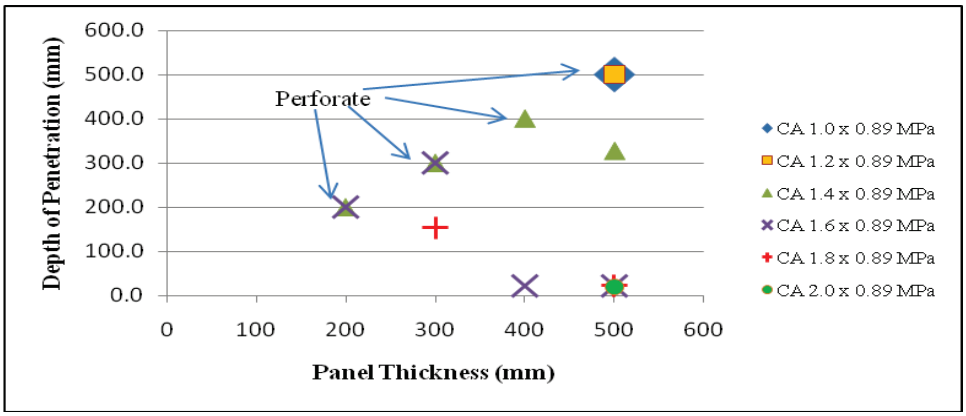



**Figure 6: Depth of penetration vs thickness of rubber panel.**

## 4. CONCLUSION

In this study, numerical simulations on the feasibility of rubber materials as bullet stoppers have been successfully conducted using the MSC Dytran software. The Mooney-Rivilin hyperelastic strain energy model was used during the simulation to represent the rubber's failure behaviour. Different values of CA were considered for this study. It was found that depth of penetration reduces with increasing CA. Based on the results, the chosen material parameters of rubber to be used in the design of the new panel are CA of 1.4 x 0.89 MPa, with thickness of 500 mm.

The simulation results demonstrate that rubber is a potential material to be used as bullet stoppers. However, the results show that rubber alone as a material for the panel is not sufficient. Indeed, the required panel geometry (500 mm in thickness) will be too large to be used in practice. Moreover, producing such massive rubber panels would be difficult. Therefore, further work should be conducted, such as in the design of appropriate geometry and configuration of panel properties in order to reduce the thickness of the panel.

## REFERENCES


Barauskas, R., Kuprys, M. (2005). Collision Detection and Response of Yarns In Computational Models of Woven Structures. *Proc. 10th Int. Conf. MMA2005&CMAM2*, pp. 3-13.

Børvik, T., Olovsson, L., Dey, S. & Langseth, M. (2011). Normal and oblique impact of small arms bullets on AA6082-T4 aluminium protective plates. *Int. J. Impact Eng*., **38**: 577-589.

Dean, G. & Read, B. (2001). Modelling the behaviour of plastics for design under impact. *Polym. Test*., **20**: 677-683.

Dorogoy, A., Rittel, D. & Brill, A. (2010). A study of inclined impact in polymethylmethacrylate plates. *Int. J. Impact Eng*., **37**: 285-294.

Duan, Y., Keefe, M., Bogetti, T.A., Cheeseman, B.A. & Powers, B. (2006). A numerical investigation of the influence of friction on energy absorption by a high-strength fabric subjected to ballistic impact. *Int. J. Impact Eng*., **32**: 1299-1312.

Gama, B.A. & Gillespie, J.W. (2011). Finite element modeling of impact, damage evolution and penetration of thick-section composites. *Int. J. Impact Eng*., **38**: 181-197.

Iqbal, M.A., Chakrabarti, A., Beniwal, S. & Gupta, N.K. (2010). 3D numerical simulations of sharp nosed projectile impact on ductile targets. *Int. J. Impact Eng*., **37**: 185-195.

Kishore (2010). *Rubber Cultivation in Malaysia*. Available online at: http://www.kish.in/rubber_cultivation_in_malaysia (Last access date: 18 November 2011).

Krishnan, K., Sockalingam, S., Bansal, S. & Rajan, S.D. (2010). Numerical simulation of ceramic composite armor subjected to ballistic impact. *Compos. Part B - Eng*., **41**: 583-593.

Lamberts, A.P.T.M.J. (2007). *Numerical Simulation of Ballistic Impacts on Ceramic Material*. Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven.

Logan, D.L. (2002). *A First Course in the Finite Element Method 3rd Ed.* Brooks / Cole, Pacific Grove, California.

MSC (2006). *MSC Dytran 2005 r3 Theory Manual*. MSC Software Corporation, Santa Ana, California.

Nyström, U. & Gylltoft, K. (2011). Comparative numerical studies of projectile impacts on plain and steel-fibre reinforced concrete. *Int. J. Impact Eng*., **38**: 95-105.

Ong, C.W., Boey, C.W., Hixson, R.S. & Sinibaldi, J.O. (2011). Advanced layered personnel armor. *Int. J. Impact Eng*., **38**: 369-383.

Qasim, H.S. (2009). Impact resistance of a rectangular polycarbonate armor plate subjected to single and multiple impacts. *Int. J. Impact Eng*., **36**: 1128-1135.

Roberts, B.J. & Benzies, J.B. (1977). The relationship between uniaxial and equibiaxial fatigue in gum and carbon black filled vulcanizates. *Proc. Rubbercon '77*, vol. 2.1, pp. 2.1–2.13.

Sands, J.M., Fountzoulas, C.G., Gilde, G.A. & Patel, P.J. (2009). Modelling transparent ceramics to improve military armour. *J. Eur. Ceram. Soc*., **29**: 261-266.

Sareen, A.K. & Smith, M.R., (1996). Evaluation of an Analytical Design Tool for Ballistic Dynamics Simulation. *American Helicopter Society 52nd Forum*, pp. 1271-1282.

Sauer, M. (2011). Simulation of high velocity impact in fluid-filled containers using finite elements with adaptive coupling to smoothed particle hydrodynamics. *Int. J. Impact Eng*., **38**: 511-520.

Savvateev, A.F, Budin, A.V, Kolikov, V.A & Rutberg, P.H.G (2001). High-Speed penetration into sand. *Int. J. Impact Eng*., **26**: 675-681

Sombatsompop, N. (1998). Practical use of the Mooney-Rivlin Equation for determination of degree of crosslinking of swollen NR vulcanisates. *J. Sci. Soc. Thailand*., **24**: 199-204.

Gander, T.J. & Cutshaw, Q.C. (Eds.) (2000). *Jane's Ammunition Handbook 2000 – 2001*. Jane's Information Group Ltd., Coulsdon, Surrey.

Wang, Y., Miao, Y., Swenson, D., Cheeseman, B.A. & Yen, C.F. (2010). Digital element approach for simulating impact and penetration of textiles. *Int. J. Impact Eng*., **37**: 552-560.

Zeng, H. B., Pattofatto, S., Zhao, H., Girard, Y. & Fascio, V. (2010). Perforation of sandwich plates with graded hollow sphere cores under impact loading. *Int. J. Impact Eng*., **37**: 1083-1091.

# FINITE ELEMENT ANALYSIS (FEA) OF A C130 TOWING BAR

Fadzli Ibrahim* & Mohammad Shafiq Toha

Mechanical & Aerospace Technology Division (BTJA), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

Email: fadzli.ibrahim@stride.gov.my

## ABSTRACT

*This paper provides a case study on the structural integrity of a towing bar using finite element analysis (FEA). The overall design of the towing bar in terms of shape and material properties was subject to the original equipment manufacturer's (OEM) design for the C130 aircraft. The simulation analysis was done for four different cases, where force was applied at four different angles; 0, 30, 45 and 60˚. The results of the FEA simulations show that the maximum force that can be applied to the towing bar reduced drastically when the load is applied at increasing angles. The computed maximum load for angle of˚,60 without taking into consideration the safety factor, is 15.85 kN, which is 2.9 times greater than the value specified in the aircraft's towing procedure. This investigation only shows the maximum force that can be applied to the specific towing bar and the behaviour of its structure during failure. Therefore, further experimental analysis should be conducted in order to reaffirm the findings from the simulation analysis.*

**Keywords**: *Towing bar; finite element analysis (FEA); maximum stress; maximum deflection; towing angle.*

## 1. INTRODUCTION

When a towing vehicle taxis a C130 aircraft, a towing bar is connected between the aircraft and vehicle. One end of the towing bar is fixed to the front wheel of the aircraft, while the other end is attached to the back of the vehicle. During operation, the towing bar is subjected to high stresses to its structure due to the high load applied. The allowable angle for turning during the towing operation is from 0˚ to 60˚ (Manson, 2011). Due to the high stress applied, towing bars can fail during operation. Therefore, it is necessary that this stress condition should be analysed to verify the claim.

Static or dynamic loads cause stresses in materials and structures that they are applied to. The analysis of these stresses is an important engineering discipline

called stress analysis, which is required for the study and design of structures, such as mechanical parts and structural frames, under given or expected loads. For structures that do not yet exist, the study of stress analysis can also be applied for its design step (Hibler, 2005).

Finite element analysis (FEA) has become more important in recent years. Numerical solutions even for a very complicated stress problem can be obtained using this method (Roylance, 2001). The model of a design or material can be computerised for a FEA system to analyse for specific stress related results. The suggested design can be verified using FEA, to determine whether it can perform under the required specifications prior to construction or manufacturing. The solid bodies of the material are discretised by FEA into small and finite volumes called the finite elements, where elasticity principles can be easily applied. Its procedure requires a pre-processing operation that converts a computer aided design (CAD) model into a discretised form of mesh. A solver, containing various equations, then evaluates the mesh. Finally, a post-processing module interprets the results obtained from the solver (Zienkiewicz *et al*., 2005; Rao, 2010).

This paper provides a case study of a structural behaviour of the towing bar of a C130 aircraft using FEA. The objectives to conduct this investigation are:

a) To determine the maximum forces that can be applied to the towing bar at critical angles that have been identified for future reference of users;

b) To analyse the behaviour of the towing bar structure towards each loading conditions.

## 2. METHODOLOGY

There are two main parts in this investigation, modelling the structure using CAD and analysing the structure using computer aided engineering (CAE). Using the dimensions given, the design of the towing bar is modelled using a CAD software, NX 6.0. The generated model of the towing bar is then analysed using a CAE analysis software, MSC Patran-Nastran.

MSC Patran is a commonly used pre- / post-processing software that provides solid modelling and meshing. It also provides analysis setup tools that streamline the creation of analysis ready models for MSC Nastran, which is an FEA solver that ~~is~~ designed to simulate stress, dynamic loadings or vibration of complex systems. The integration of MSC Patran and MSC Nastran software provides sufficiently accurate results to evaluate the performance of products (MSC, 2012).

## 2.1 Design

Figure 1 shows the model of the towing bar designed using NX 6.0. The structure is designed to withstand a C130 aircraft weighing at 54,000 kg (critical weight) at the angle from 0-60˚ at the horizontal plane (Manson, 2011). A boundary condition is identified as a fixed end with all the degrees of freedom being constrained at the flat end of the structure (Jafari *et al*., 2006).

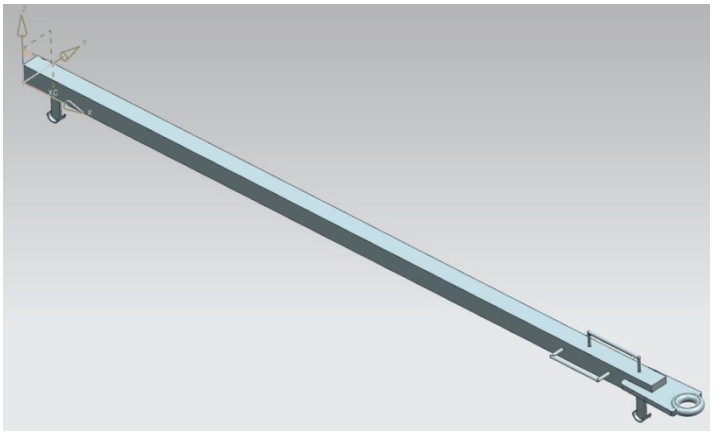


**Figure 1: Design of the towing bar.**

In order to ensure that the towing bar shall withstand the given pressure, the material used to manufacture this structure is AISI 1040 steel (Mac Donald, 2007; Shigley & Mischke, 1989), with the properties shown in Table 1.

**Table 1: Material properties of AISI 1040 steel.**
**(Source: George *et al.*, 2002; Hibler, 2005)**

| Properties | Value |
|---|---|
| Modulus of elasticity $E$ | 201 GPa |
| Poisson ratio $v$ | 0.29 |
| Maximum yield strength | 447 MPa |
| Ultimate tensile strength | 640 MPa |

## 2.2 FEA Simulations

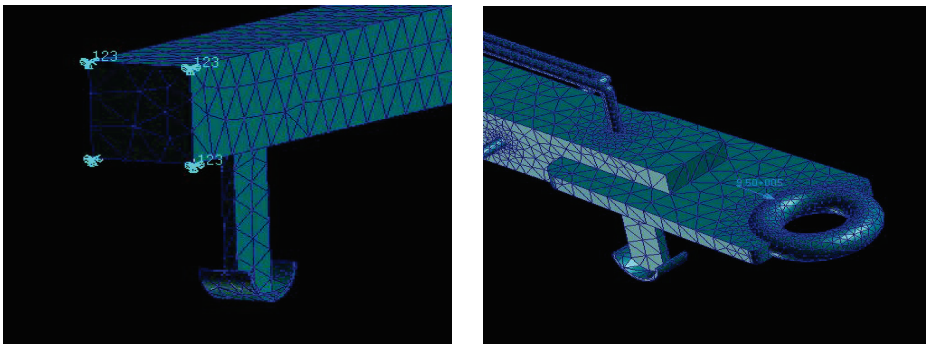
The model of the structure is then imported into MSC Patran-Nastran for its structural analysis. The mesh is created using the software and the boundary conditions are determined to simulate the real life situation that the structure

experiences during its operation (Figure 2). There are four cases that are found to be practical for this analysis; the load at angles of 0, 30, 45 and 60. For each case, the load is increased from 10 kN up to a point where the maximum stress experienced by the structure reaches the maximum yield strength.

The yield function of material deformation is the distortion-energy theory and is called the Von-Mises theory. Therefore, it is the most suitable theory to be used for ductile materials (Shigley & Mischke, 1989). Von-Mises stress $\sigma'$ is calculated using the following equation:

$$\sigma' = \left[\left\{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2\right\}/2\right]^{1/2} \quad (1)$$

where $\sigma_1$, $\sigma_2$ and $\sigma_3$ are principal stresses associated with the three principal directions (Jafari *et al*., 2006).




**(a)** **(b)**

**Figure 2: Meshed and boundary condition of the towing bar created using MSC Patran-Nastran: (a) fixed to the towing vehicle; (b) force by the aircraft.**

## 3. RESULTS AND DISCUSSION

The parameters that are recorded for each case are the load when the structure yields and the maximum deflection of the structure. Figure 3 shows the results of the FEA simulations conducted on the towing bar model for loadings at different angles. These results show the maximum force and deflection that the structure can withstand at the respective angles of loading.

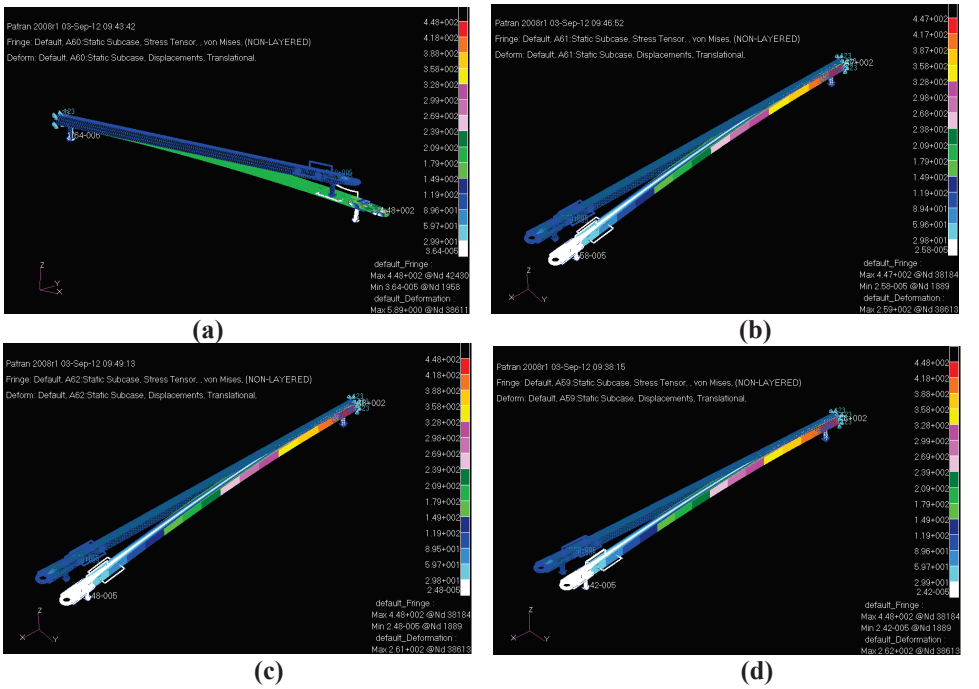



**Figure 3: FEA simulation results for different angles: (a) 0˚; (b) 30˚; (c) 45˚; and (d) 60˚.**

Figure 4 shows the region where the maximum stress occurs and the position where the maximum deflection takes place when the load applied at a certain angle $\theta$. For the loading condition at angle 0 the maximum stress occurs at the area of the pin, and the maximum deflection takes place at the horizontal plane (elongation of the structure).

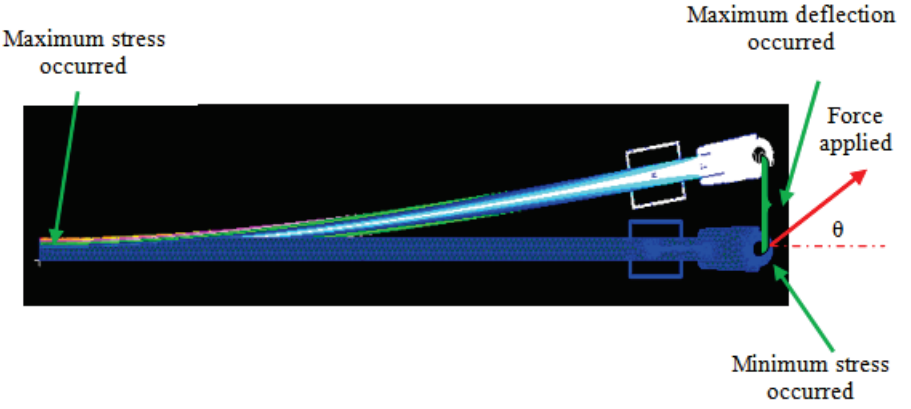



**Figure 4: Locations where maximum stress and deflection occur when force is applied at a certain angle $\theta$.**

Since the critical weight of the aircraft during operation is roughly 54,000kg (5.4kN), the maximum force would have to be greater for the structure not to fail. In a situation where the load applied to the towing bar is greater than the maximum force for a certain angle, the structure would deform permanently or even break (Xu *et al*., 2007). Table 2 shows the maximum forces that the structure can withstand at the studied angles of load and the maximum deflection before it fails at maximum stress (maximum yield strength of 447MPa).

**Table 2: Maximum force and maximum deflection at different angle before the structure yielded.**

| Degree (°) | Maximum Force (kN) | Maximum Deflection (mm) |
|---|---|---|
| 0 | 950 | 5.89 |
| 30 | 27.2 | 259 |
| 45 | 19.35 | 261 |
| 60 | 15.85 | 262 |

Based on the results obtained from the analysis, maximum force drastically reduces when the load is applied at increasing angles. On the other hand, the maximum deflection of the structure increases, indicating that the greater the angle of the applied load, the easier it is for the structure to deform or fail (Suripa & Chaikittiratana, 2008). This follows that stress increment always associates with a given strain increment (Mingzhou *et al*., 2002). According to the towing procedures, for safety reasons, the maximum weight of the aircraft during towing at the maximum angle of operation of 60 is 5.4 kN (Manson, 2011). However, the results show that at the maximum operating angle of 60˚ (without considering safety factor) the maximum load that can be applied is 15.85 kN, which is 2.9 times greater than the allowable load. This finding shows that there is a significant difference between the two, and it is almost impossible for the towing bar to fail during its operation, taken into consideration that its use follows the standard procedures provided.

## 4. CONCLUSION

In this study, FEA simulations of a C130 towing bar under static loading was carried out. The maximum forces that can be applied to the towing bar reduced drastically when the load is applied at increasing angles, while the maximum deflection of the structure increases. The computed maximum load for angle of 60˚, without taking into consideration the safety factor, is 15.85 kN, which is 2.9 times greater than the value specified in the aircraft's towing procedure. This analysis provides users with acceptable data for the operation of the towing bar at different angles. The simulation results for the deflections also give qualitative

(visual) and quantitative (maximum deflection) data on how the towing bar's structure behaves in the case of failure.

The FEA simulations conducted in this study only identified the maximum force as well as the locations of maximum stress and deflection when the load is applied to the towing bar. It is suggested that further experimental analysis be conducted in order to reaffirm the findings of this study.

## REFERENCES


George, S.B., Henry, R.C. & John, A.V. (2002). *Materials Handbook, 15th Ed*., McGraw Hill Companies, New York.

Hibler, R.C. (2005). *Mechanic of Materials*. Prentice Hall, New York.

Jafari, A., Khanali, M., Mobli, H. & Rajabipour, A. (2006). Stress analysis of front axle of JD 955 combine harvester under static loading. *J. Agri. Soc. Sci.,* **2**: 133-135.

Mac Donald, B.J. (2007). *Practical Stress Analysis with Finite Elements.* Glasnevin, Dublin.

Manson (2011). *Maintenance Instruction on Ground Handling, Service and Aircraft Maintenance.* Manson Co. Inc*.,* Tucson, Arizona.

Mingzhou, S., Qiang, G. & Bing, G. (2002). Finite element analysis of steel members under cyclic loading. *Finite Elements in Analysis and Design***.** **39**: 43–54

MSC (2012). *Patran : FE Modeling and Pre/Post Processing*. Available online at: http://www.mscsoftware.com/Products/CAE-Tools/Patran.aspx (Last access date: 18 September 2012)

Rao, P.N. (2010). *CAD/CAM Principles and Applications, 3rd Ed.* McGraw Hill, New Delhi.

Roylance, D. (2001). *Finite Element Analysis*. Massachusetts Institute of Technology, Cambridge.

Shigley, J.E. & Mischke, C.R. (1989). *Mechanical Engineering Design, 5th Ed.* McGraw Hill Companies, Singapore.

Suripa, U. & Chaikittiratana, A. (2008). Finite element stress and strain analysis of a solid tyre. *J. Achieve. Mater. Manuf. Eng.*, **31**: 576-579

Xu, C.G., Liu, G.H., Ren, G.S., Shen, Z., Ma, C.P. & Ren, W.W. (2007). Finite element analysis of axial feed bar rolling. *Acta Metall. Sin. (Engl. Lett.)*, **20**: 463-468

# FAILURE ANALYSIS OF A MARINE VESSEL SHAFT COUPLING

Nik Hassanuddin Nik Yusoff[1*], Mahdi Che Isa[1], Mohd Moesli Muhammad[1], Hasril Nain[1], Mohd Subhi Din Yati[1], Syed Rosli Sayd Bakar[2] & Irwan Mohd Noor[1]

[1]Marine Materials Research Group, Maritime Technology Division (BTM)
[2]Mechanical & Aerospace Technology Division (BTJA)
Science and Technology Research Institute for Defence (STRIDE),
Ministry of Defence, Malaysia

[*]Email: nhassanuddin.nyusoff@stride.gov.my

## ABSTRACT

*This paper presents a failure analysis conducted for a shaft coupling used in a ship's propulsion system, which failed while running at low speed during a sea trials assessment. The fracture occurred at the connection between the cylindrical parts and flange. A detailed metallurgical investigation was carried out to determine the failure initiations and failure mode. The failure was dominated by fatigue due to the appearance of ratchet marks, which were as a result of multiple fatigue cracks propagation, linking each other. Based on the marks' direction, it was determined that the failure was influenced by low load and high stress concentrations of torsion. The initial failure point was found at the keyway throughout the beach marks present. Higher magnification of cracks region found that the crack growth was induced by pitting corrosion.*

**Keyword**: *Failure analysis; fatigue; shaft coupling; pitting corrosion; stress localisation.*

## 1. INTRODUCTION

Fatigue is a damage process caused by the growing of cracks due to cyclic stress that generates and aggregates microcracks which can cause sudden catastrophic failures. In practice, 90% of all mechanical failures are due to fatigue, which occurs under repeated application of stress which is too small to cause failure in a single application in the elastic region (Fonte & Freitas, 2009).

A shaft coupling is a part of the propeller shaft system that is located next to the gearbox system to align between the intermediate and gearbox shafts. This component is subjected to fluctuating loads of combined bending and torsion with various degrees of stress concentrations. Any defects that occur on the surface or internal structure of the shaft coupling will expose it to fatigue failure

(ASM, 1986, 1992, 1993; Pan *et al.*, 2006). Failures of such components and structures have engaged scientists and engineers extensively to determine their main causes, and thereby, propose methods for their possible prevention. In general, such shafts run with a steady torsion load superimposed with bending stress either due to the shaft's self-weight or possible misalignment between the bearings. In some cases, such shafts also suffer from corrosion fatigue. The chances of the component failing due to fatigue are higher from the combined effect of both corrosion and repeated stresses acting together than those factors acting separately (Du *et al.*, 1998).

This paper presents a failure analysis conducted for a shaft coupling used in a ship's propulsion system, which failed while running at low speed during a sea trials assessment. The fracture occurred at the connection between the cylindrical parts and flange. A detailed metallurgical investigation was carried out to determine the cause of failure of the shaft coupling.

## 2. METHODOLOGY

Before the failed shaft coupling was examined, it was ultrasonically cleaned with acetone for 30 min in order to remove particles of fracture residue and to clearly reveal the fracture marks. Then it was inspected visually, and then macro- and microscopically. Macroscopic examination was performed using a Stemi DV4 stereo microscope, while microscopic observation was conducted using a Carl Zeiss Axiovert 25 inverted metallurgical light microscope equipped with a couple charged device (CCD) camera. Both images were analysed with the aid of the Axio Vision KS 400 software.

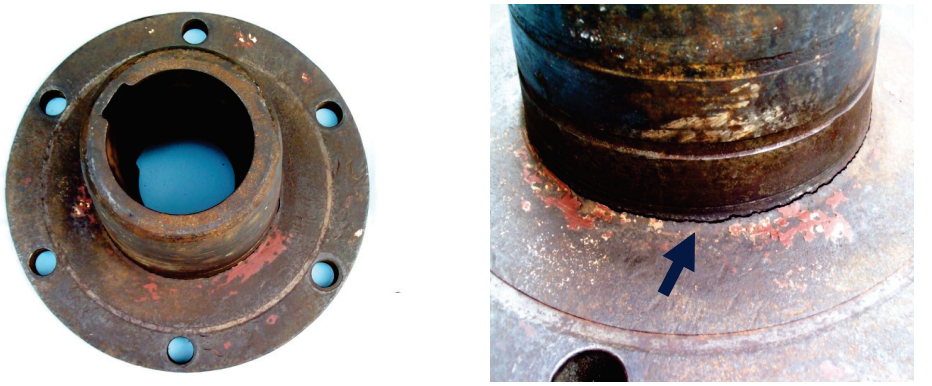
The chemical compositions of the shaft coupling were determined using a Bruker S4 Pioneer wave dispersive X-ray (WDXRF), while the hardness of the shaft coupling was measured at ten different locations using a Rockwell Affri DRMC 250 hardness tester.

The shaft coupling was subjected to detailed metallographic examination in order to collect additional information about the internal features on the microscopic level. The samples were cut, mounted, grinded and polished according to metallurgical protocol to reveal the internal microstructure of the alloy. Microstructure observations were performed using a Carl Zeiss Axiovert 25 inverted metallurgical light microscope

## 3. RESULTS

### 3.1 Visual Examination

Figure 1 shows the general appearance of the shaft coupling and its failure spot. Visual examination of the shaft coupling showed that the residual of anticorrosive coating still remained on the shaft coupling. The location of the crack was at the connection between cylindrical parts with the flange. The shaft coupling was disassembled to its cylindrical and flange parts, as shown in Figure 2, to inspect the fracture surface.



(a) (b)

**Figure 1: General appearance of the (a) failed shaft coupling and (b) its failure spot.**


(a) (b)

**Figure 2: The shaft coupling disassembled to (a) cylindrical and (b) flange parts.**

From the visual examination of the failure surface, the multiple fatigue cracks origin sites on the flange part of the shaft coupling are shown in Figure 3. Figure 4 shows the magnified view of the crack surface zone. Figure 4(a) shows the fracture surface where there are quite obvious ratchet marks that occurred as a result of linking of the propagation among several fatigue cracks. This is

indicating the boundary of two adjacent failure planes. Figure 4(b) shows river marks which look like rivers (Neville, 2005) that occurred as a result of fast-growing sections of the fatigue zone. These marks also indicate the direction of the cracks' growth. The overall appearance of the examined crack surface indicates that the fracture occurred due to the influence of torsion stress. The combination of many ratchet marks and a small overload zone, as shown in Figure 3, indicates that the load was low but there were high stress concentrations, as discussed by Neville (2005).

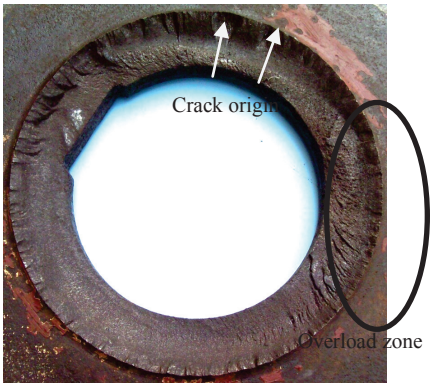
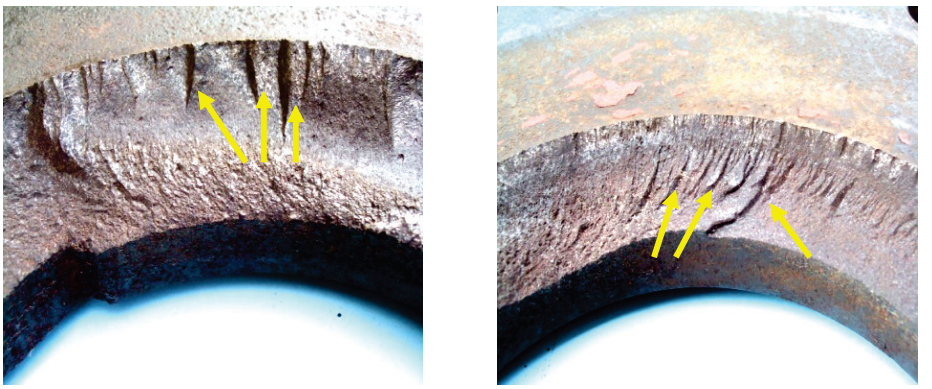



**Figure 3: The multiple fatigue cracks origin sites on the flange part of the shaft coupling.**

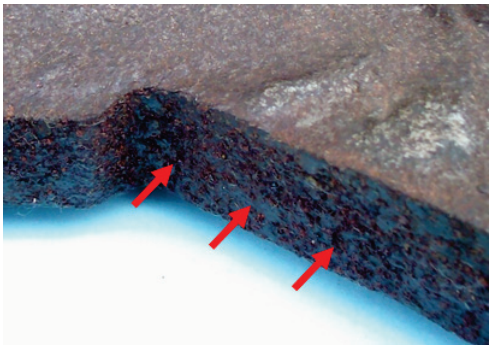


**(a)** **(b)**

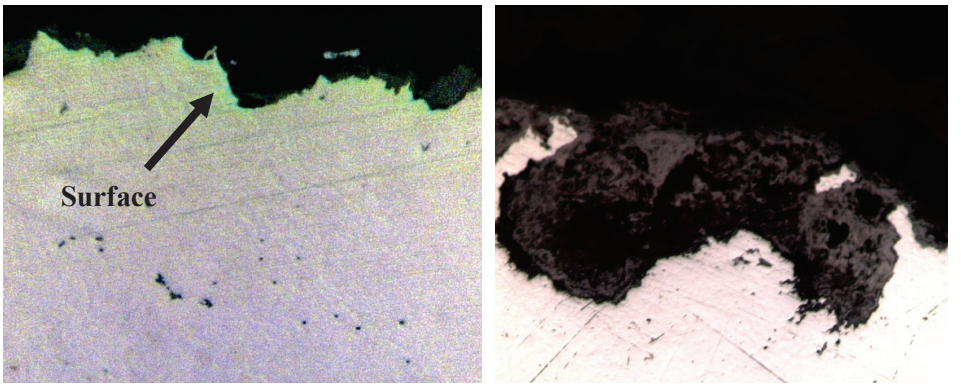
**Figure 4: Magnified view of the crack surface zone that consists of (a) ratchet and (b) river marks.**

### 3.2 Metallographic Inspection

Figure 5 shows a part of the shaft coupling where pitting corrosion occurred on the fracture surface. Examination on the cross section image and its magnification (Figure 6) found that pitting corrosion was clearly present on the fracture surface that caused an uneven surface. This type of corrosion occurs on the open surfaces when the protective layers either passive film (that naturally protects the metal surface) or protective coating breakdown (Frankel, 1998). The exposure of the shaft coupling to the marine environment that containing the chloride ions causing the breakdown of the passive film and initiation of pitting corrosion (Du, 1998).



**Figures 5: A part of the shaft coupling where pitting corrosion (indicated by the arrows) occurred close to the fracture surface.**




**(a)** **(b)**

**Figure 6: Examination of the fracture surface: (a) The cross section image. (b) High magnification of the cross section.**

Figure 7 shows the close up image of the fracture surface. It also observed lines known as beachmarks (shown by the white arrows) which correspond to the changes of load in services. These line is helpful to calculate the number of

cycles among different changes of load which allows the reconstruction of the history of the crack propagation (Fonte & Freitas, 2009). The presence of the beachmarks was found on the surface of the keyway. It can be seen that the fatigue originated at the centre of beachmarks from one of the pits on the fracture surface as show by the yellow arrow.

Figure 8(a) shows the microstructure of the shaft coupling which consists of the white region which is known as ferrite and dark region as pearlite structure. Ferrite is a essentially pure carbon in the form of body centered cubic structure. While pearlite is a mixture of ferrite and cementite. It occurs from the transformation of austenite on slow cooling (Maranian, 2010). The presence of this structure can be seen clearly in the high magnification image shown in Figure 8(b), the dendrite structures cannot be found from the micrograph. This proves that the shaft coupling had gone through proper heat treatment process to break the dendrite structure that is formed from the casting process (Momcilovic *et al.*, 2012).

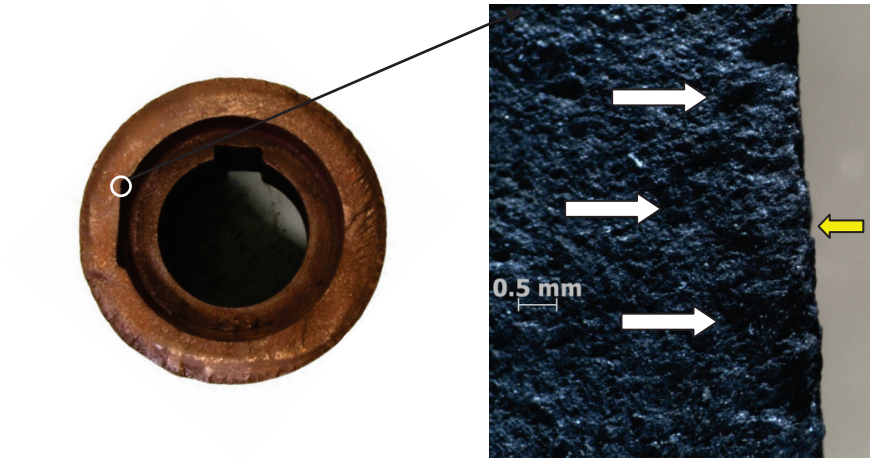
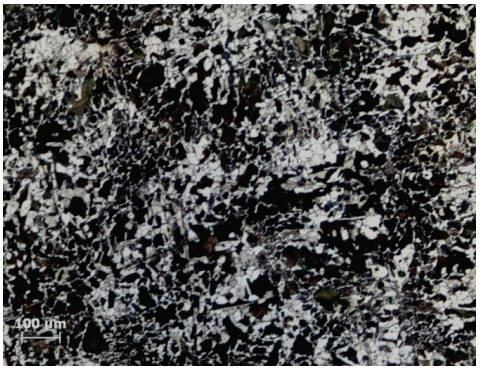
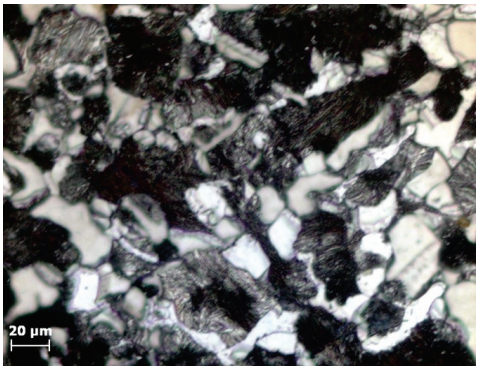



**Figure 7 : Fatigue beachmarks present on the fracture surface. Note that the origin of the fatigue is a pit shown by yellow arrow. The white arrows show beach marks present on the surface.**




**(a)**




**(b)**

**Figure 8: The microstructure of the shaft coupling with (a) low magnification and (b) high magnification images.**

### 3.3 Material Properties

The results of the material composition investigation are summarised in Table 1. According to Radzikowska (2004), the higher content of carbon, in the range between 2 to 4 %, in the steel alloy shows that the shaft coupling is cast iron. Cast iron, in general, is an engineering material that is widely used in pipes, machines, shafts and engine manufacturing parts, such as cylinder heads and blocks, and gearbox cases. It is relatively castable, and has excellent machinability, resistance to deformation and wear resistance (Smith, 1996). Other elements such silicon (Si), manganese (Mn), sulphur (S) and aluminium (Al) have specific functions to improve the mechanical properties of the cast iron (Thisza, 2001).

**Table 1: Chemical compositions of the failed shaft coupling obtained via WDXRF.**

| Element (%) | | | | | |
|---|---|---|---|---|---|
| Carbon | Si | Mn | S | Al | Ferum |
| 2.1 | 0.59 | 0.94 | 0.065 | 0.008 | 96.297 |

Hardness values were measured using the Rockwell hardness scale B on the cross section of the failed shaft coupling. The average hardness value obtained is 53.7 HRB.

## 4. DISCUSSION

It is necessary to address the following points in order to analyse the failure of the coupling shaft:

(a) Whether the material of the coupling was defective
(b) Whether the operating conditions had any influence on the degradation and failure of the coupling.

The material is as per the chemical composition and the microstructure of the white alloyed cast iron (Radzikowska, 2004). There is neither inhomogenity nor microstructure degradation of the coupling shaft. Therefore, the possibility of material failure due to material degradation is ruled out.

It was determined by visual inspection that the anticorrosive coating was delaminated from the metal surface. Exposure to corrosive environments without any protective coating allowed corrosion to attack the shaft coupling. As can be seen in the Figure 5, pitting corrosion can be clearly found on the surface as a result of atmospheric corrosion. The atmospheric corrosion rate depends on relative humidity, and the presence of various gases and solid particles in the air. In high humidity, chloride compounds that are found in marine atmospheres increase the rate of atmospheric attacks (ASM, 2005). The pitting produced is a localised corrosion attack which leads to the formation of small pits concentrated in the same area, before forming bigger holes in the corrosion resistant metal (Askeland *et al.*, 2006). As a result, pitting corrosion could be detrimental because it can affect material properties by decreasing its strength performance (Jones, 2001).

The crack growth found on the fracture surface shows that the failure began from multiple points. The progression marks found indicate that the fracture occurred due to torsion failure with low load but high stress concentrations. The initial point of failure had been contributed by stress at the keyway as indicated by the beach marks found close to the keyway. This is in good agreement with Neville (2005), who stated that keyways are exposed to the highest stress concentrations for components.

## 5. CONCLUSION

In order to identify the cause of the shaft coupling's failure, the investigation conducted involved visual, metallurgical inspection and mechanical properties of the sample. Visual inspection found that an inappropriate anticorrosive coating on the metal surface had allowed the corrosion attack. As a result, pitting could be seen clearly on the shaft coupling's surface, which had demolished its strength. The crack growth found on the fracture surface occurred due to torsion force with high stress concentrations. The initial point of failure had been contributed by stress at the keyway, suggesting that the failure was due to fatigue corrosion.

## ACKNOWLEDGEMENTS

The authors would like to thank the officers and staff from various laboratories at STRIDE Kajang and Lumut for their technical support during the investigation works.

## REFERENCES

American Society for Metals (ASM) (1986). *ASM Metals Handbook, Volume 11: Failure Analysis and Prevention*. American Society for Metals (ASM), Russell Township, Ohio.

American Society for Metals (ASM) (1992). *ASM Handbook of Case Histories in Failure Analysis, Vol. 1 and Vol. 2.* American Society for Metals (ASM), Russell Township, Ohio.

American Society for Metals (ASM) (1993). *ASM Handbook of Case Histories in Failure Analysis, Vol. 2.* American Society for Metals (ASM), Russell Township, Ohio.

Askeland, D.R. & Fulay, P.P. (2006). The Science and Engineering of Materials, Bill Stenquist, Australia.

Cramer, S.D. & Covino, B.S (2005). *ASM Metals Handbook, Volume 13B: Corrosion: Materials*, ASM International, United State of America.

Du, M.L., Chiang, F.P., Kagwade, S.V. & Clayton, C.R. (1998). Damage of Al 2024 alloy due to sequential exposure to fatigue, corrosion and fatigue. *Int. J. Fatig.*, **20**: 743-748.

Fonte, M. & Freitas, M.D. (2009). Marine main engine crankshaft failure analysis: A case study. *Eng. Fail. Anal.*, **16**: 1940-1947.

Frankel, G.S. (1998). Pitting Corrosion of Metals A Review of the Critical Factors. *J. Electrochem. Soc.,* **145**: 2186-2198.

Jones, D.R.H. (2001). *Corrosion of Central Heating Systems, in Failure Analysis Case Studies II*. Pergamon, Oxford.

Maranian, P. (2010). *Reducing Brittle and Fatigue Failures in Steel Structures*, ASCE Publications, Virginia.

Momcilovic, D., Odanovic, Z., Mitrovic, R., Atanasovska, I. & Vuherer, T. (2012). Failure analysis of hydraulic turbine shaft. *Eng. Fail. Anal.*, **20**: 54-66.

Neville, N.M. (2005). Understanding the Surface Features of Fatigue Fractures: How They Describe the Failure Cause and the Failure History. *J. Fail. Anal. Prevent.*, **5**: 11-15.

Pan, H.L., Tang, S.H. & Hao, J.W. (2006). Failure analysis of a rotating cantilever shaft in chloride corrosive environment. *Eng. Fail. Anal.*, **13**(4): 646-655.

Radzikowska, J.M. (2004). Metallography and microstructures of cast iron. *In* Vander Voort, G.F. (Ed,), *ASM Handbook: Metallography and Microstructures*. ASM International, Ohio, pp. 565-587.

Smith, W.F. (1996). *Principles of Material Science Engineering*, McGraw-Hill, New York.

Thisza, M. (2001). *Physical Metallurgy for Engineers*. ASM International and Freud Publishing House Ltd, Ohio USA.

# THE EFFECT OF AIR GAP THICKNESS ON SOUND ABSORPTION COEFFICIENT OF POLYURETHANE FOAM

Mohd Moesli Muhammad*, Noor Aishah Sa'at, Hasril Naim, Mahdi Che Isa, Nik Hassanuddin Nik Yussof & Mohd Subhi Din Yati.

Marine Materials Research Group, Maritime Technology Division (BTM), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*Email: moesli.muhammad@stride.gov.my

## ABSTRACT

*Polyurethane foam is widely used in noise control engineering to absorb sound. This paper investigates the effect of air gap thickness behind polyurethane foam on the sound absorption coefficient for low and high frequency sounds using the impedance tube method. The polyurethane foam test samples were prepared with thickness of 25 mm, with two different diameters; 29 and 100 mm for high (1.0 to 6.4 kHz) and low (100 Hz to 1.6 kHz) frequency measurements respectively. The foam was subjected to microscopic observation under an optical microscope for pore size analysis. It was observed that the sample has a honeycomb structure with the majority of pores' diameters being less than 1 mm. The measurement was carried out for six different air gap thicknesses; 0, 5, 10, 15, 20 and 25 mm. The results showed that introducing an air gap behind the sample influences the sound absorption coefficient, which increased especially for higher frequency sounds until the optimum value was obtained. It also showed that the frequency of maximum peaks for varying air gap thicknesses was different, with the peaks for larger air gap thicknesses shifting towards lower frequencies. However, this combination of polyurethane foam and air gap was not able to absorb low frequency sounds, especially below than 250 Hz.*

**Keyword**: *Acoustical material; polyurethane foam; air gap thickness; sound absorption coefficient; impedance tube.*

## 1. INTRODUCTION

The applications of noise control are at present given significant priority in various industries, such as automotive, manufacturing and ship building. It plays an important role in creating an acoustically pleasant environment. This can be achieved when the intensity of sound is reduced to a certain level that is not harmful to human ears. Various techniques can be applied for this, which employ

different kinds of materials. One such technique is using acoustical materials to absorb sound (Gracia-Valles *et al*., 2008; Yang & Wu, 2011; Jaouen & Becot, 2011). These acoustical materials are available in the market as fibrous or porous materials, and in various types, such as nonwovens, fibrous glass, mineral wools and foams (Ersoy & Kucuk, 2009; Arenas & Crocker, 2010; Kino & Ueno, 2007). Generally, all these types of materials are made from polymeric or rubber based materials, due to low cost of production and their flexibility, making it easy to cut and form into complex shapes. However, for acoustical materials in high temperature applications, such jet engines, most of them are made from ceramics that are hard and fire resistant (Zhang *et al*., 2006; Cuiyun *et al*., 2012; Fuji *et al*., 2006).

The main function of acoustical materials is to reduce the acoustic energy of sound waves that pass through it (Figure 1). This can be performed using resistive materials that consist of porous structure which change sound energy into heat. This can happen when energy is changed due to frictional forces between sound waves and the cell walls in the pore structure (Fang *et al*., 2007). In other words, the further the distance sound waves travel through the medium of porous structure, the higher the amount of sound energy which is dissipated. From the reduction of sound energy, it can be assumed that this energy is being absorbed by the acoustical materials. Therefore, the amount of sound energy that is absorbed represents the sound absorption characteristics which describe performance of the acoustical materials.

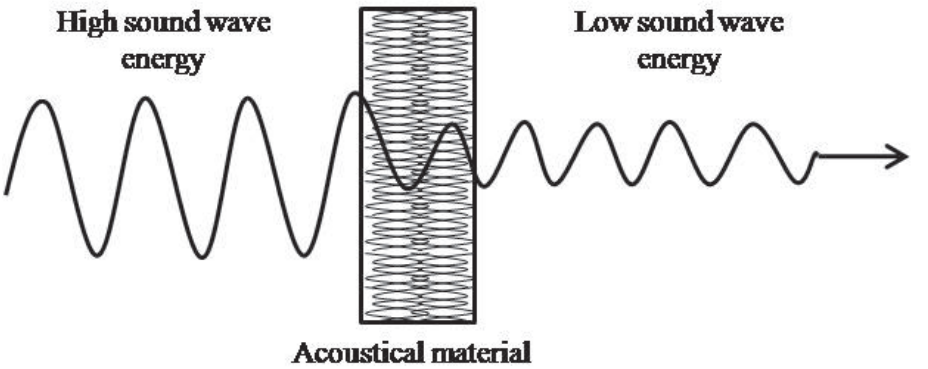



**Figure 1: High sound wave energy is reduced after passing through the acoustical material.**
**(Source: Bies & Hansen, 2009)**

In practice, there are two methods used to install acoustical materials; either using an air gap or not. It can be attached directly, or some air gap thickness can be created from the wall. The methods of installing acoustical materials play an

important role in improve noise absorption. By introducing an air gap behind an acoustical material any residues of incident sound waves that are not absorbed or transformed into heat energy after transmission through the acoustical material will face additional resistance through the air medium. The mechanism of sound waves dissipating in the air medium is known as the Helmoltz resonance effect. It excites some of the sound waves' frequencies, causing them to oscillate at greater amplitude. When the energy reaches the maximum level, the sound waves become weak due to friction with air particles, which converts sound energy into heat (Zhang *et al*., 2012; Norton & Karszub, 2003; Crocker, 2007).

Previous studies show that the combination of acoustical materials and air gap thickness has significant impacts on the sound absorption coefficient. Rosli *et al*. (2009) reported that the air gap layers within the coir fibre sound absorption panels improve the sound absorption coefficient for medium and high frequencies. Seddeq (2009) found that the sound absorption coefficient increases when plastic fibre absorbers are associated with air gap. The advantage of installing acoustical materials with air gaps is to reduce costs by maintaining the thickness of the acoustical materials while improving the sound absorption coefficient.

This study is aimed at investigating the effect of air gap thickness on the sound absorption coefficient of polyurethane foam. It is a porous type of acoustical material which has been widely used in many applications, including construction, automotive, machinery and electronics. Even though polyurethane foam is excellent for noise control applications, not much attention has been provided on the combination of polyurethane foam with air gaps. The findings of this study demonstrate that adding an air gap behind polyurethane foam improves it sound absorption coefficient for high frequencies applications.

## 2. METHODS FOR EVALUATING THE PERFORMANCE OF ACOUSTICAL MATERIALS

There are two different methods available to evaluate the performance of acoustical materials, which are reverberant room and impedance tube (Sagartzazu *et al*., 2007; Ingard, 2010; Doutres, 2010). In general, the measurement is to study the effects of exposure of materials to known sound fields. Both methods are able to collect the properties of sound absorption materials, such as sound absorption coefficient, reflection coefficient and surface impedance. Sound absorption coefficient is the property that is most referred to by engineers and scientists in evaluating acoustical environments (Crocker, 2007). By looking at the value of the sound absorption coefficient, the ability of acoustical materials to absorb the sound energy can be predicted, and it is used in calculation at early stage of design. The reverberant room method needs expensive measuring equipment, a qualified acoustician and a large size of samples (Scien, 2011; Bies

& Hansen, 2009). Therefore, this method is limited and cannot be expanded upon due to the high costs involved in setting up the facility and preparing the samples.

In the impedance tube method, the test sample is tested within a rigid tube, in which the sound is internally guided, forced to propagate along the tube's axis and hits the back plate (Crocker, 2007). Figure 2 shows the schematics of the impedance tube method. In general, the main components of the impedance tube are a loudspeaker, two microphones of 0.25 in and back plate. The loudspeaker placed at one of the tube end creates a sinusoidal pressure disturbance that propagates down the tube with the test sample positioned at the other end of tube near the back plate. The back plate's function is to reflect the incident waves and to hold the sample in the proper position. Its position is unfixed and is moveable if the measurement is carried out for air gap environments. The two microphones are in fixed positions and both of them measure simultaneously the incident wave before the sound wave enters the test sample, and reflected wave after the sound waves propagates through the test sample, hits the back plate and travels back towards the loudspeaker. The signals of incident and reflected waves from the microphones are analysed using a signal analyser based on the transfer function method. Using this method, the ratio of reflected and incident waves is obtained, which represents the sound absorption coefficient of the acoustical material (Doutres, 2010).

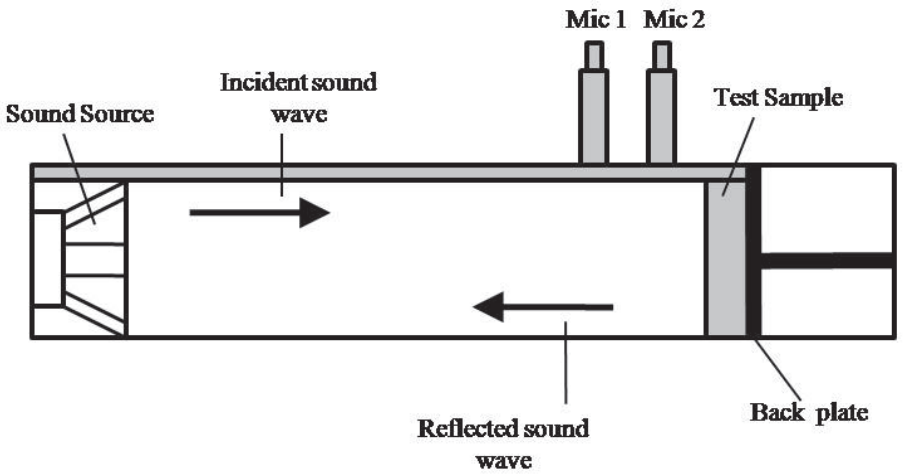



**Figure 2: Schematic of the impedance tube method.**
**(Source: Crocker, 2007)**

Most of the measurements for sound absorption coefficient are carried out using two tubes of different diameters; large and small diameters for low and high frequency sound respectively. Both measurement and design of the impedance tubes are according to the international standards, which are ISO (1998) and JIS

(2007). According to Crocker (2007), to ensure that only plane waves, with no transverse waves, enter tube, the lengths of the tubes shall exceed $0.25\lambda$, while the diameters shall not exceed $0.58\lambda$, where $\lambda$ is the wavelength of sound in air. For example, tubes with diameter of 100 mm and length of 910 mm are useful for sounds in the range 90 Hz to 1.8 kHz. In order to conduct the measurement over the range of 90 Hz to 6.0 kHz, two different tube sizes are required. Therefore, many of the impedance tube products available in the market have a large tube of 100 mm diameter for low frequency sounds and a small tube of 29 mm diameter for high frequency sounds. Based on the diameters of the tubes, the frequency ranges are 100 Hz to 1.6 kHz and 1.0 to 6.4 kHz for low and high frequency sounds respectively. However, as the frequency range between 1.0 to 1.6 kHz is repeated for both the low and high impedance tubes, variations of the results may exist. Therefore, in order to increase accuracy, many researchers suggest that the values of sound absorption coefficient be computed based on the averages of the measured data.

In the impedance tube, which is a close boundary system, the calculations are made based on the assumption that the propagation of the waves is in 1D direction. Therefore, only plane waves propagate in the tube. The calculations of the sound absorption coefficient using the impedance tube method have been established and described in many publications (Crocker, 2007; Hansen, 2009). The calculations are quite complex and are based on the transfer function method. To simplify these calculations, the sound absorption coefficient $\alpha$ is the ratio between intensities of reflected and incident waves, and can be expressed as:

$$\alpha = 1 - \frac{I_r}{I_i} \tag{1}$$

where $I_r$ and $I_i$ are the intensities of the reflected and incident waves respectively. If the ratio uses the spectrum of fast Fourier transform (FFT), it can be expressed as:

$$\alpha = 1 - \frac{S_{bb}}{S_{aa}} \tag{2}$$

where $S_{bb}$ and $S_{aa}$ are the FFT spectrums of the reflected and incident waves respectively.

The value of $\alpha$ is usually expressed in the range of 0 and 1. A material that absorbs all incident waves will have $\alpha$ = 1. On the other hand, if $\alpha$ = 0, no energy of sound is absorbed. Therefore, this material also can be called an

insulation material and if it is backed with a rigid wall, all the incident waves will be fully reflected by the material (Crocker, 2007).

## 3. METHODOLOGY

The test apparatus was part of a complete acoustical material testing system from Scien Co. (Figure 3). In this system, a loudspeaker placed at one end of tube generates a broadband random signal from 100 Hz to 6.4 kHz. The sample was placed at the end of the impedance tube in front of the back plate. Two fixed microphones were located vertically on the tube, with the distances from the loudspeaker being 150 and 170 mm for microphones 1 and 2 respectively. The microphones were of the 0.25 in free-field type from SIEN Co., which were used to measure incident and reflected sound waves in the impedance tube respectively. For data acquisition and signal processing, an embedded power amplifier, a four-channel SCIEN Co Vibro-Acoustic ADC 3241 signal platform and a desktop computer equipped with Acoustic Duct Version 9291-4.3E software were used.

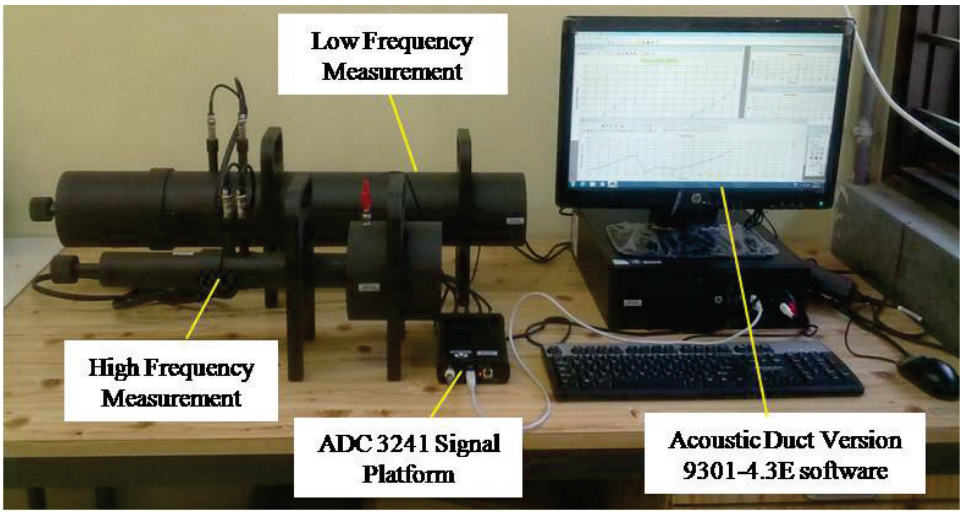



**Figure 3: The Scien Co. impedance tube measurement system was used in this study.**

The polyurethane foam sample was subjected to macroscopic observation under a Carl Zeiss Stemi DV4 optical stereo microscope and the image was analysed using the Axio Vision KS 400 software. For the sound absorption coefficient measurement, test samples were prepared in two different diameters, 29 and 100 mm for high and low frequency measurements respectively (Figure 4). The

thickness of samples was fixed at 25 mm. For the measurement of sound absorption coefficient with air gap thickness, the study was carried for six different air gap thicknesses, which were 0, 5, 10, 15, 20 and 25 mm. Figure 5 shows the setup of air gap thickness, which is the separation between the back plate and test sample that creates a cavity.

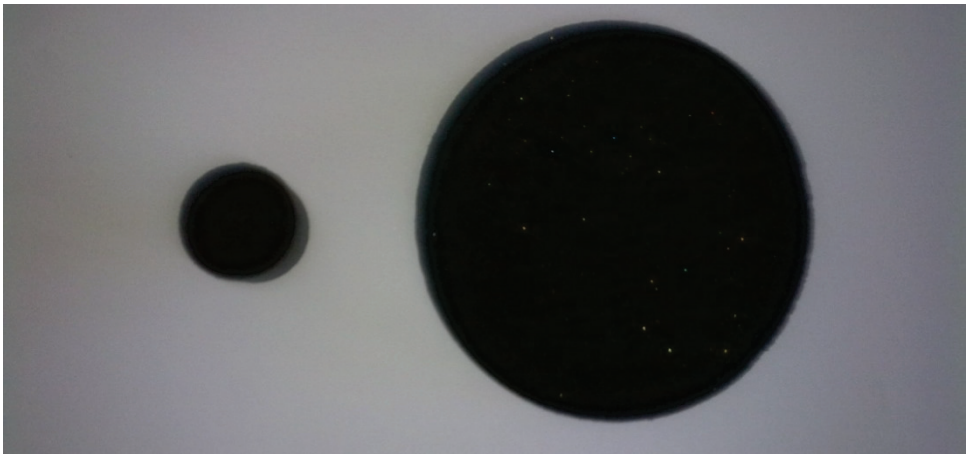


**Figure 4: The prepared of polyurethane foam test samples with two different diameters; 29 mm for the small foam on the left and 100 mm for the large foam on the right.**

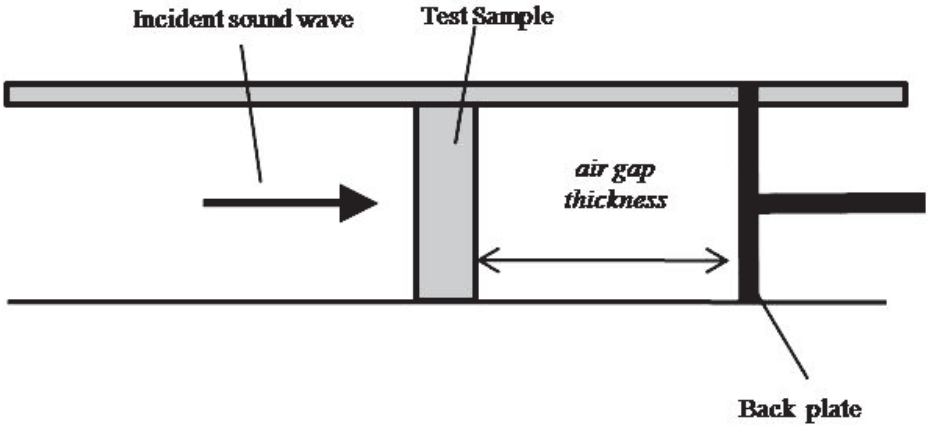



**Figure 5: The separation between the back plate and test sample that creates a cavity which represents air gap thickness.**

## 4. RESULTS & DISCUSSION

The observation using the optical microscope (Figure 6) shows that the structure of the polyurethane foam sample is pore cells with honeycomb structure. The pore cells are interconnected to each other with multiple pore cell sizes. The bright areas in the image correspond to the pores, whereas the dark areas correspond to the open holes. Detailed examination of the image shows that the majority of the pores' diameters are less than 1 mm.

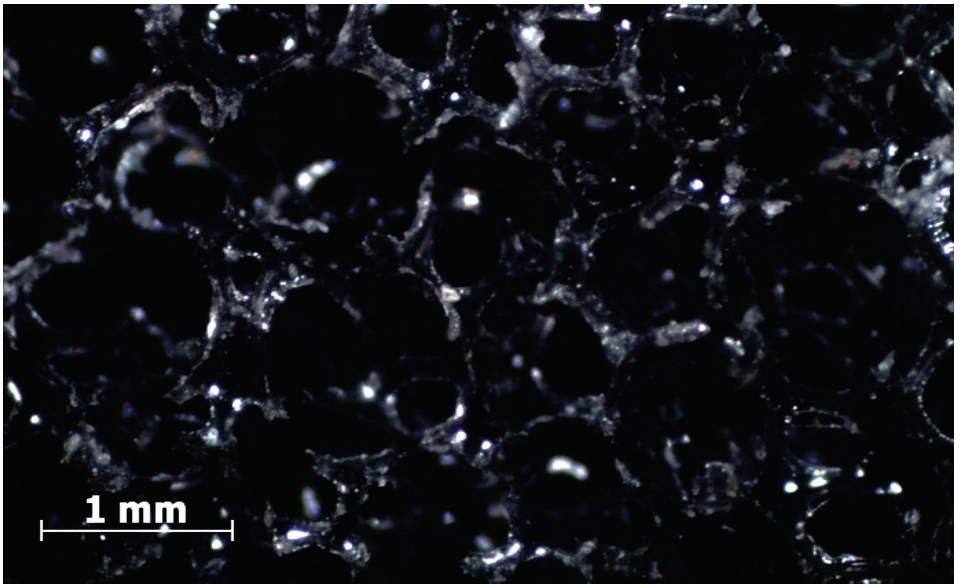



**Figure 6: Image of the porous polyurethane foam captured using an optical microscope.**

Figure 7 shows the correlation between sound absorption coefficient and wave frequency for the different air gap thicknesses. It can be seen that the properties of sound absorbing characteristics of polyurethane foam vary significantly with wave frequency. All the plots show the same trend, which is low sound absorption coefficients at low frequencies, and high sound absorption coefficients at high frequencies.

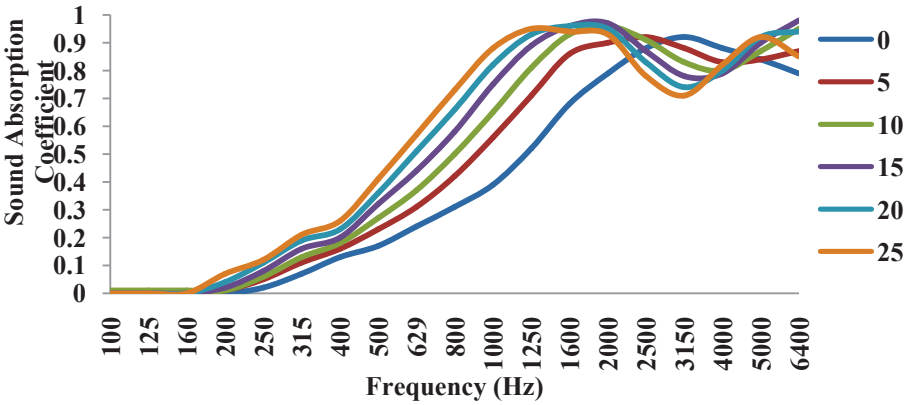



**Figure 7: Sound absorption coefficients of polyurethane foam with different air gap thicknesses.**

The plot for polyurethane foam without air gap shows the maximum peak value of sound absorption coefficient of 0.92 at frequency of 3.15 kHz. As the frequency is increased, the values of absorption decrease. For the case of polyurethane foam with different of air gap thicknesses, the plots show two different peaks which are maximum and minimum respectively. The maximum peaks occur at medium frequencies, while minimum peaks occur at higher frequencies. As compared with the peak of the plot without an air gap, all the peaks of the plots with air gaps are shifted towards to lower frequencies with higher thicknesses of air gap showing lower values. The value of maximum peak of each plot does not show much difference with most of the peaks being close to 1, with the range of 0.05. This low range indicates that the sound absorption coefficient is not affected by air gap thickness. The highest value obtained is 0.97 for air gap thickness of 15 mm. The minimum peaks for all the plots with air gaps occur at high frequencies, with the lowest value being 0.78 for air gap thickness of 25 mm. All the plots also show that the absorption coefficient value drops between frequencies of 2 to 3.15 kHz after reaching the maximum peak, except for the sample with air gap thickness of 25 mm, where there is another drop at frequency of 5 kHz.

For the analysis of sound absorption coefficients of the polyurethane foam with and without air gap, the values obtained indicate that both methods are efficient in absorbing sound waves at high frequencies, particularly above 250 Hz. However, for the polyurethane foam with air gaps, all the plots show the improvements of the sound wave frequencies that are being absorbed. For very low frequencies, below 250 Hz, the results show that sound absorption coefficient with value of zero is obtained. This indicates that both methods are not able to absorb the incident waves and fully reflected after hitting the back plate. According to Zhang *et al*. (2012), low frequency sounds are very difficult

to absorb because of their long wavelength. As a result, the total of sound waves energy remaining constant. This phenomenon is in good agreement with the principle of conservation of energy (Giordano, 2010). Therefore, in any isolated or closed system, the sum of all forms of energy remains constant. It can be concluded that the polyurethane foam with thickness of 25 mm, with and without air gaps used in this study are unable to absorb low frequency sound waves, and hence, is inefficient for low frequencies applications. In order to improve the reduction of low frequency sounds, the acoustical materials used should be thicker in order to provide enough time for sound waves to form into heat when travelling through the obstacles within the acoustical materials (Seddeq, 2009; Norton *et al*., 2003).

The mechanism of sound waves dissipated in air medium is due to the Helmoltz resonance effect (Zhang *et al.*, 2012; Norton *et al*., 2003; Crocker, 2007). The peaks of resonance frequencies can be clearly seen in the Figure 7, which correspond to the maximum sound absorption coefficient at these frequencies. However, the sound absorption coefficient at certain frequencies is reduced after reaching the maximum peak. This phenomenon occurs due to sound waves propagation within the air medium, where the excitation and degradation of sound energy depend on the thickness of air gaps (Ayub *et al.*, 2009; Zhang & Tianning, 2009; Fouladi *et al.*, 2010). Even though the sound absorption coefficient is reduced, the values obtained are still high, with most of them above than 0.8.

## 5. CONCLUSION

Microscopic observation showed that the polyurethane foam acoustical material used in this study is porous with honeycomb structure. The results of the analysis showed that the majority of diameters of the pores are less than 1 mm. The sound absorption coefficient of polyurethane foam with six different air gap thicknesses of 0, 5, 10, 15, 20, and 25 mm were studied and validated through experimental measurements in an impedance tube measurement system. Sound absorption coefficient for both with and without air gaps significantly increased with increase of sound wave frequencies, until the maximum peak was obtained. All the maximum peaks of the plotted of air gap thicknesses shifted with the lowest frequency obtained for larger air gaps. The results also showed that the air gaps reduced the sound absorption capability at high frequencies of sound waves. Both methods, with and without air gap, were not able to absorb the low frequencies of sound waves, particularly below than 250 Hz. The implementation of combination polyurethane foam with air gaps was able to enhance the sound absorption coefficient in the medium frequency range without changing the thickness of acoustical materials.

## ACKNOWLEDGEMENTS

This study was conducted as part of the Tenth Malaysian Plan (RMK10) project entitled *A Study of Vibroacoustic Properties of Composite Materials*. The authors would like to thank the Science and Technology Research Institute for Defence (STRIDE) for providing research facilities and technical assistance.

## REFERENCES

Arenas, P. J. & Crocker, J. M. (2010). Recent trends in porous sound absorbing materials. *Sound Vib.*, **43**: 12-16.

Ayub, Md., Mohd. Nor, M.,J., Amin, N. & Zulkifli, R. (2009). A preliminary study of effect of air gap on sound absorption of natural coir fiber. *Proceedings of the Regional Engineering Postgraduate Conference, 20-21 October 2009*. National University of Malaysia, Bangi, Malaysia.

Bies, D. A. & Hansen C. H. (2009). *Engineering Noise Control.* Spon Press, New York.

Crocker, M. J. (2007). *Handbook of Noise and Vibration Control.* John Wiley & Sons, Inc, New Jersey.

Cuiyun, D., Guang, C., Xinbang, X. & Peisheng L. (2012). Sound absorption characteristics of a high-temperature sintering porous ceramic material. *Appl. Acoust.*, **73**: 865-871.

Doustress, O., Salissou, Y., Atalla, N. & Panneton, R. (2010). Evaluation of the acoustic and non-acoustic properties of sound absorbing materials using a three –microphone impedance tube. *Appl. Acoust.*, **71**: 506-509.

Ersoy, S. & Kucuk, H. (2009). Investiagation of industrial tea-leaf-fibre waste material for its sound absorption properties. *Appl. Acoust.*, **70**: 215-220.

Fang, W., Lu-cai, W., Jian-guo, W. & Xiao-hong, Y. (2007). Sound absorption property of open-pore aluminium foams. *China Foundry*, **4**: 31-33.

Fouladi, M. H., Mohd Nor, M. J., Ayub, Md. & Leman, Z. A. (2010). Utilization of coirfiber in multilayer acoustic ab sorption panel. *Appl. Acoust.*, **71**: 241-249.

Fuji, M., Kato, T., Zhang, F. & Takahashi, M. (2006). Effects of surfactants on the microstructure and some intrinsic properties of porous building ceramics fabricated by gel casting. *Ceram. Int.*, **32**: 797-802.

Giordano, N. J. (2010). *College Physics: Reasoning & Relationships*. Cengage Learning, Independence, Kentucky.

Gracia-valles, M., Avilla, G., Martinez, S., Terradas, R. & Nogues, J.M. (2008). Acoustic barriers obtained from industrial waste. *Chemosphere*, **72**: 1098-1102.

Ingard, U. (2010). *Noise Reduction Analysis.* Jones & Barlett Publisher, United Kingdom.

ISO (International Standard Organization) (1998). *Acoustics - Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes - Part 2:*

*Transfer Function Method*. International Standard Organization (ISO), Geneva.

JIS (Japan International Standard) (2007). *Acoustics - Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes - Part 1: Method Using Standing Wave Ratio.* Japan International Standard, Japan.

Jayaraman, K. A. (2005). *Acoustical Absorptive Properties of Nonwovens.* Masters thesis, North Carolina State University, Raleigh, North Carolina.

Jaoeun, L & Becot, F. X. (2011). Acoustical characterization of perforated facings. *J Acoust. Soc. Am.*, **129**: 1400-1406.

Kino, N. & Ueno T. (2007). Investigation of sample size effects in impedance tube measurements. *Appl. Acoust.*, **68**: 1485-1493.

Norton, M.P. & Karszub, D.G. (2003). *Fundamental of noise and vibration analysis for engineers.* Cambridge University Press, United Kingdom.

Rosli, Z., Mohd Jailani, M.N., Ahmad Rasdan, I., Mohd Zaki, N. & Mohd Faizal, M.T. (2009). Effect of perforated size and air gap thickness on acoustic properties of coir fibre sound absorption panels. *Eur. J. Sci. Res.*, **28**: 242-252.

Sagartzazu, X., Hervella, L. & Pagalday, J. M. (2007). Review in sound absorbing materials. *Arch. Comput. Method E*, **15**: 311-342.

Scien (2011). *Acoustic Duct, User's Manual*. Scien Co. Ltd., South of Korea.

Seddeq, H.S. (2009). Factors influencing acoustic performance of sound absorptive materials. *Aust. J. Basic Appl. Sci.,* **3**: 4610 -4617.

Wang, C. N. & Torng, J. H. (2001). Experimental study of the absorption characteristics of some porous fibrous materials. *Appl. Acoust.*, **62**: 447-459.

Yang, S. & Yu W. D. (2011). Air permeability and acoustic absorbing behaviour of nonwovens. *J. Fiber Bioeng. Inf.,* **3**: 204-208.

Zhang, F, Z., Kato, T., Fuji, M. & Takahashi, M. (2006). Gelcasting fabrication of porous ceramics using a continuous process. *J. Eur. Ceram. Soc.*, **26**: 667-671.

Zhang, B. & Tianning, C. (2009). Calculation of sound absorption characteristics of porous sintered fiber metal. *Appl. Acoust.*, **70**: 337-346.

Zhang, C., Li, J., Hu, Z., Zhu, F. & Huang Y. (2012). Correlation between the acoustic and porous cell morphology of polyurethane foam: Effect of interconnected porosity. *Mater. Design*, **41**: 319-325.

## CATCH OF THE NET

Photonics is the science and technology of generating, transmitting, amplifying, modulating and detecting light, and its applications for practical purposes. It builds heavily on optical technology, supplemented with modern developments such as lasers, amplifiers, light-emitting diodes (LEDs), optical fibres, optical modulators and photodetectors. Photonics technology development and applications have substantially increased across over the past several years. The applications of photonics are ubiquitous, including in fields such as information technology (optical fibre communications and optical data storage), health care (medical diagnostics and therapy in ophthalmology, and laser surgery), metrology (time and frequency measurements, and rangefinding), industrial manufacturing (laser material processing, semiconductor chip manufacturing), entertainment (laser shows and holographic art), and defence and space technology (satellite surveillance systems, navigation, night vision imagers, missile guidance, anti-missile systems and high-power directed-energy weapons). With expected future applications including energy conversion and photonic quantum computing, photonics has become established as an enabling technology for a multitude of industries and many areas of economy. The following are relatively interesting and useful websites on photonics and their applications:

1) **RP Photonics: Encyclopedia of Laser Physics and Technology**
   http://www.rp-photonics.com/encyclopedia.html

   An encyclopaedia providing explanations on various concepts and fields of photonics.

2) **Institute of Physics (IOP): Optics and Photonics - Physics Enhancing Our Lives**
   http://www.iop.org/publications/iop/2009/page_38205.html

3) **National Academy of Science: Optics and Photonics - Essential Technologies for Our Nation**
   http://www.nap.edu/catalog.php?record_id=13491

   Reports discussing the applications of photonics in various fields, and potential future technologies and applications.

4) **Photonics.com**
   http://photonics.com

5) **Photonics Online**
   http://www.photonicsonline.com

6) **Optics & Photonics News (OPN)**
   http://www.osa-opn.org

   Resource centres providing articles, materials and up-to-date news on photonics technologies, applications and issues.

7) **Boston University Photonics Center**
   http://www.bu.edu/photonics

8) **Stanford Photonics Research Center**
   http://photonics.stanford.edu

9) **Dublin Institute Of Technology: Photonics Research Centre**
   http://aoc.dit.ie

10) **Universiti Teknologi Malaysia (UTM): Photonics Technology Centre (PTC)**
    http://www.fke.utm.my/ptc

    Research institutes working on the development of photonics technologies in various fields of applications, including autonomous vehicle navigation, integrated circuits, quantum information processing, microscopy, nanophotonics, neuroscience and solar cells.