JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

Table of Contents

Table of Contents

Evaluation of a Virtual Prototype of the First Saudi Arabian Car

Mustufa H. Abidi, Ali Ahmad, A. M. El-Tamimi, S.M. Darwish, A. M. Al-Ahmari
King Saud University
mustufahaider@yahoo.com

Abstract

Recurring, efficient, and widespread use of prototypes can differentiate between successful and unsuccessful entry of new products into a competitive marketplace. Physical prototyping can be a lengthy and costly process, particularly if alterations resulting from design evaluation involve tool redesign. Advanced computer technology enabled the utilization of digital prototypes, i.e. they are virtual rather than physical. Virtual prototypes witnessed great success in large automotive and aerospace industries. It is important to assess how representative is the developed virtual prototype when compared to the real world counterpart, and the sense of presence reported by users of the virtual prototype. This paper evaluates the representativeness and the sense of presence for a virtual prototype of the first Saudi Arabian designed car. The virtual prototype is evaluated for various aesthetics and design features in a semi-immersive virtual environment. The results of the user-based evaluation indicated that the car virtual prototype is representative of the real Saudi Arabian car and offers a flexible environment to study design features when compared to its physical prototype.

1. Introduction

Computer-Aided Engineering (CAE) is a popular technique that may reduce products design's cycle times and limit the need of costly prototypes [1]. During product development, many design and assembly decision issues need to be solved [2]. Prototypes can provide solutions to such issues and hence influence the product development process. The accessibility and affordability of advanced computer technology has paved the way for increasing utilization of virtual instead of physical prototypes. Virtual Prototyping (VP) is about presentation and analysis of three-dimensional Computer-Aided Design (CAD) models before creating any physical prototypes [1].

Advances in computing speeds and enhanced models' representativeness of physical phenomena, along with the growing ability to move results between various types of modeling and analysis environments resulted in improving the scope of applications, robustness, accuracy, realism, and cost effectiveness of VP [3]. VP offers several capabilities such as the development and viewing of three-dimensional solid models with various colors and surface textures. It allows for animation of mechanisms, finite element analysis, ergonomics analysis, crash testing for automobile bodies, etc. [4-8]. Commercially available VP tools offer the following functionalities [1]:

a) Mechanical design, e.g. two/three-dimensional drafting, sketching and solid modeling
b) Shape design and styling to deal with innovative forms and complex shapes

c) Analysis and simulation solutions including stress analysis, design optimization in terms of mass, displacement and principle stresses, and kinematic and dynamic simulation.

VP builds on top of virtual reality (VR) technology, which immerses a user in the same virtual environment as that of model. Jasnoch et al. suggest that VP consists of three domains: virtual environment, simulation, and manufacturing process design [9]. On the other hand, Pratt states that almost any form of a computer model will serve for some purpose as a VP [10]. Presently, most commercially-available VP applications consider VR an integral part, due to the fact that VR offers a sense of presence in the virtual environment, which can augment the user's capabilities in terms of product evaluation. Examples for VP use include automotive design validation [4], rapid product development [5], haptics evaluation of devices with knobs [6], assembly analysis of cable harness [7], and design and assembly of connecting rods interactively [8]; among many others.

Prototypes are typically used for communication, design development, and design verification [11]. Zorriassatine et al. identified five broad classes of VP based on their purposes and goals [1]:

1. Visualization
2. Fit and interference check of assemblies
3. Testing and verification of functions
4. Evaluation of operations
5. Human factors analysis

This paper discusses the development and evaluation of a VP of the first Saudi Arabian designed car. The VP was built for visualizing car design decisions such as interior and exterior color choices, evaluate the ergonomics, and fit and interference check of mechanical assembly.

2. Virtual prototyping for visualizing models

Visualization models are used for assessment of form, shape, and appearance [2]. They play a vital role in communicating product information among a diverse group of users including advertising agents, customers, managers, product development teams, engineering, and repair and maintenance operators. In addition, visual appearance may serve as an attraction factor [1]. Baxter explained three aspects of product attractiveness that are articulated through product appearance [12]:

1. Attraction due to recognition of previously used products
2. Symbolic attraction (appeal to personal and social values of customers)
3. Intrinsic attraction (inherent beauty of the product form)

Presently, most commercial CAD software offer three-dimensional solid model's photo pragmatic still and dynamic images that fulfill all appearance necessities. Visual acuity and appeal can be assessed for digital models with a variety of forms concerning product layout, illustrating the overall product components and subassemblies, their associated colors, and surface textures and finishes. On the other hand, VP provides lively three-dimensional viewing from any angle, in addition to graphical animations that can be utilized to represent a variety of conditions and scenarios depending on the target audience. New visualization software has the potential to simulate interactive navigation such as variable-speed walking and flying through complicated assemblies of any size; thus facilitating efficient and realistic visual inspection [13].

3. Virtual prototyping for fits and interference check

It is important to check how well subassemblies or individual components fit with the rest of parts. Since it is cost-ineffective and sometimes impossible to manufacture parts with accurate dimensions and features, dimensional tolerances are included in product design to accommodate variations due to manufacturing processes [2]. Fits and interference checks at an early stage of product design may reduce scrap, defects, and rework; thus reducing the total cost of a product. Traditional fits and interference checks are time-consuming and error-prone [14]. VP provides novel mechanisms for fits and interference checks by providing realistic three dimensional models and automated real time collision detection; which results in a listing of all the interferences, where clearances and interfering regions of the model can be highlighted using different colors [15]. Figure 1 shows an interference check on the Saudi Arabian car prototype.

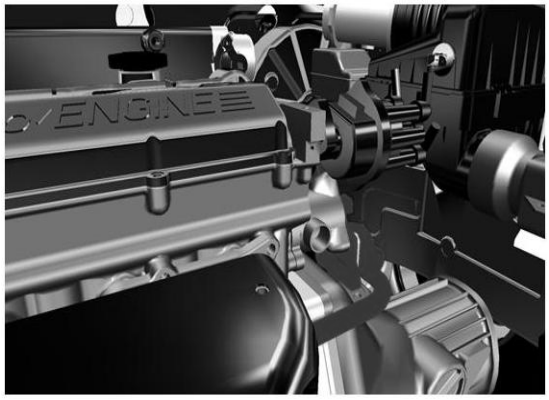


Figure 1. Real Time Interference Check

Visual inspection of a digital three-dimensional assembly can be accomplished using a wide range of options such as dynamic viewing of any part from any angle, or via virtual flights throughout the assembly, where magnified viewing permits close assessment. It is also possible to validate the ability of sliding or mating parts to obtain intended movements at each level of dimensional values within tolerance zones [1].

4. Development of virtual prototype of the first Saudi Arabian designed car

The name of the first Saudi Arabian designed passenger car is Gazal-1. It is developed by professors, students, engineers and technicians at the Advanced Manufacturing Institute in King Saud University, Riyadh, Saudi Arabia. Gazal-1 is a sport utility vehicle (SUV), based on a Mercedes-Benz G-Wagon platform. Gazal-1 is 4.8 m long and about 1.9 m wide and named after a desert deer. A VP is developed to analyze and review Gazal-1's design, which enables exploring various aesthetics features like different colors, lighting, conditioning, material, etc. in an interactive semi-immersive virtual environment. Figure 2 shows the Gazal-1 VP.

Figure 2. Virtual Exploration of Car Model

5. Evaluation of virtual prototype of the first Saudi Arabian car

This paper presents the results of assessing the utility of the developed VP, and identifying opportunities for improvement, where a modified presence questionnaire is used to capture users' feedback [16]. The questionnaire consists of twenty five questions on a scale of 1 to 5, where 1 indicates very poor, 2 indicates

poor, 3 indicates average, 4 indicates good, and 5 indicates very good.

5.1. Participants

Sixteen male students from King Saud University (mean age 23.94 years; S.D. = 1.06) participated in this study. All participants reported normal or corrected to normal vision. All the participants were right handed.

5.2. Apparatus

The experimental setup consists of Dell precision computer used to generate graphics and integrates various hardware systems. A Christie Mirage projector is used for rear end projection on a 3.1m X 2.3m screen to create a semi-immersive virtual environment. Active stereo shutter glasses were used by the participants for stereoscopic viewing. Head and hand tracking was done using an Intersense IS-900 motion tracking system. Participants navigated and interacted with the virtual model through a hand wand.

5.3. Procedure

Prior to start of a testing session, participants went through an experimenter-led familiarization training of the various semi-immersive virtual environment devices. Then, the participants completed several operations and tasks on a 1:1 virtual prototype of Gazal-1 such as: examining the VP with different exterior colors, changing car seats' material, operating car radio, opening and closing car's doors and hood, and examining the car from various angles. Once these tasks were completed, participants filled a presence questionnaire.

5.4. Results

After completing all the tasks, the participant's subjective sense of presence was assessed using the modified presence questionnaire. Table 1 presents descriptive statistics of user's data. The overall average score on the presence questionnaire was 4.02 (S.D. = 0.5821).

Table 1. Descriptive Statistics

No	Question	Mean	S.D.
1	How much were you able to control events?	4.125	0.619
2	How responsive was the environment to actions that you initiated (or performed)?	4.250	0.683
3	How natural did your interactions with the environment seem?	3.438	0.512
4	How completely were all of your senses engaged?	3.813	0.403
5	How much did the visual aspects of the environment involve you?	4.688	0.602
6	How natural was the mechanism which controlled movement through the environment?	3.688	0.704
7	How aware were you of events occurring in the real world around you?	3.938	0.574
8	How aware were you of your display and control devices?	3.813	0.834
9	How compelling was your sense of objects moving through space?	4.188	0.655
10	How inconsistent or disconnected was the information coming from your various senses?	2.188	0.403
11	How much did your experiences in the virtual environment seem consistent with your real-world experiences?	3.813	0.655
12	Were you able to anticipate what would happen next in response to the actions that you performed?	3.313	0.479
13	How completely were you able to actively survey or search the environment using vision?	4.250	0.683
14	How well could you actively survey or search the virtual environment using touch?	2.375	0.500
15	How compelling was your sense of moving around inside the virtual environment?	4.188	0.403

16	How closely were you able to examine objects?	5.000	0.000
17	How well could you examine objects from multiple viewpoints?	4.750	0.447
18	How well could you move or manipulate objects in the virtual environment?	4.688	0.602
19	How involved were you in the virtual environment experience?	4.125	0.719
20	How distracting was the control mechanism?	3.250	0.577
21	How much delay did you experience between your actions and expected outcomes?	3.375	0.500
22	How quickly did you adjust to the virtual environment experience?	3.813	0.403
23	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	3.875	0.619
24	How much Virtual model of Gazal-1 is similar to the Real Gazal-1?	4.313	0.602
25	Is the Virtual prototype of Gazal-1 is better than the physical prototype in terms of similarity with the real one?	4.750	0.447

The results of the questionnaire analysis show that two questions resulted into means less than 2.5, which were in relation to touch feedback, and information from various senses. The absence of haptic feedback system resulted in poor environment performance with respect to force feedback. Otherwise, the virtual prototype performs satisfactorily. For the question concerned with "close examination of the object" (question 16), all users provides a rating of 5, which means that the virtual environment is suitable for the close evaluation of virtual models. Similarly, a mean of 4.75 for the question regarding the comparison of virtual prototype with the physical prototype suggests that the virtual prototype might be representative of the real car.

Figure 4 shows graphical analysis of user's feedback for selective questions.

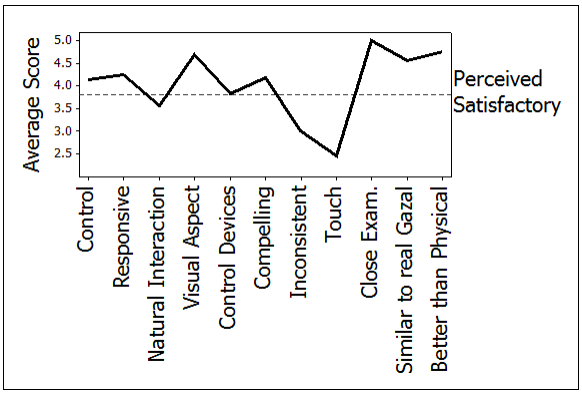


Figure 4. Users' Feedback Analysis

Based on Figure 4 it can be seen that the touch or force feedback is absent in the virtual environment. The other measures which are below the perceived satisfactory level (which is arbitrarily set at 3.8) are inconsistency, natural interaction, distraction due to control mechanism and little delay. The virtual environment provides features to view objects closely, from multiple viewpoints and it also provides the in-depth or insight view of various components. Participants concluded that the virtual prototype of Gazal-1 is similar to the real car and it is also more representative of real car than the physical prototype.

6. Discussion

Using prototypes is important for the design thinking process, but prototypes are feasible only if they provide appropriate affordances that end users can perceive and evaluate. The present study examined the utility of using immersive virtual reality for the prototyping tasks. Different design and aesthetics features of a virtual car were examined in the virtual environment. The results of the study indicated that the virtual prototype of the car is a close representative of its real model. This fact is in agreement with the previous literature [17].

One of the interesting facts that revealed out through user's feedback study is that the

developed virtual environment scores low on natural interaction measure due to the absence of touch and auditory feedback. Overall performance of the virtual prototype of car is rated better than its physical counterpart by the users.

7. Conclusions

A VP of first Saudi Arabian designed car is developed and presented in a semi-immersive virtual environment. The results showed that the virtual prototype is representative of a real car. Therefore, one can judge various design and aesthetics features of car quite effectively. With the help of collision detection technique, various interference and clashes can also be diagnosed efficiently.

On the basis of user's feedback analysis, it can be concluded that the virtual prototype of car works quite well for design review purpose. The developed virtual environment gives enough immersiveness to the user that he/she can assume the virtual world close to the real one. Descriptive statistics suggests that the virtual prototype performs well for most of the aspects. On the basis of the results, it can be concluded that virtual prototyping possesses lot of potential and it can make product design more efficient in terms of time and cost.

8. Acknowledgement

This work was supported by the Advanced Manufacturing Institute at King Saud University. The authors thank the Center of Manufacturing Technology Transfer (CMTT) at King Saud University for providing CAD models of Gazal-1.

9. References

[1] Zorriassatine F., Wykes C., Parkin R. & Gindy N. (2003). 'A survey of virtual prototyping techniques for mechanical product development'. In Proceedings of Institute of Mechanical Engineers Part B: J. Engineering Manufacture, 217, pp. 513-530.

[2] Ulrich K. & Eppinger S. (2011). 'Product Design and Development'. 5th Edition, McGraw-Hill / Irwin, ISBN-10: 0073404772.

[3] Chua C.K., Teh S.H. & Gay R.K.L. (1999). 'Rapid prototyping versus virtual prototyping in product design and manufacturing'. International Journal of Advanced Manufacturing Technology, 15, pp. 597-603.

[4] Kulkarni A., Kapoor A., Iyer M. & Kosse V. (2011). 'Virtual prototyping used as validation tool in automotive design'. In Proceedings of 19th International Congress on Modelling and Simulation, Perth, Australia, pp. 419-425.

[5] Choi S.H. & Cheung H.H. (2008). 'A versatile virtual prototyping system for rapid product development'. Computers In Industry, 59, pp. 477-488.

[6] Ha S., Kim L., Park S., Jun C. & Rho H. (2009). 'Virtual prototyping enhanced by a haptic interface'. CIRP Annals - Manufacturing Technology, 58, pp. 135-138.

[7] Sung R.C.W, Ritchie J.M., Robinson G., Day P.N., Corney J.R. & Lim T. (2009). 'Automated design process modelling and analysis using immersive virtual reality'. Computer-Aided Design, 41, pp. 1082-1094.

[8] Bourdot P., Convard T., Picon F., Ammi M., Touraine D. & Vézien J.M. (2010). 'VR-CAD integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of CAD models', Computer-Aided Design, 42, pp. 445-461.

[9] Jasnoch U., Kress H. & Rix J. (1994). 'Towards a virtual prototyping environment'. In Proceedings of IFIP WG 5.10 on Virtual Environments and Their Applications and Virtual Prototyping, pp. 173-183.

[10] Pratt M.J. (1995). 'Virtual prototypes and product models in mechanical engineering'. In Virtua Prototyping - Virtual Environments and the Product Design Process, Eds Rix J., Hass S.& Teixeira J., Chapman and Hall, London, pp. 113-128.

[11] Black R. (1996). 'Design and Manufacture - An Integrated Approach'. Macmillan Press, London.

[12] Baxter M. (1995). 'Product Design: A Practical Guide to Systematic Methods of New Product Development'. Chapman and Hall, London.

[13] Li L., Li J.S., Du F. & Si D.H. (2010). 'Building Virtual Reality Design System Based on DIVISION Mockup Software'. In proceedings of International Conference on Electrical and Control Engineering, China, pp. 552-555.

[14] Craig A.B., Sherman W.R. & Will J.D. (2009). 'Developing Virtual Reality Applications'. Morgan Kaufmann Publishers, Elsevier. ISBN: 978-0-12-374943-7.

[15] Abidi M.H., Ahmad A., El-Tamimi A.M. & Al-Ahmari A.M. (2012). 'Development and Evaluation of Virtual Assembly Trainer'. In Proceedings of Human Factors and Ergonomics Society 56th Annual Meeting, Boston, US, pp. 2560-2564.

[16] Witmer B.G. & Singer M.J. (1998). 'Measuring Presence in Virtual Environments: A Presence Questionnaire'. Presence, 7(3), pp. 225-240.

[17] Kim C., Lee C., Lehto M.R. & Yun M.H. (2011). 'Affective evaluation of user impressions using virtual product'. Human Factors and Ergonomics in Manufacturing & Service Industries, 21(1), pp. 1-13.

The Impact of Corporate Average Fuel Economy (CAFÉ) On US Auto Makers

Steve Button, Gordon W. Arbogast and Jim Mirabella
Jacksonville University
garboga@ju.edu

ABSTRACT

The Environmental Protection Agency (EPA) regulation of required U.S automaker fuel economy has been ongoing since the early 1970's. Automakers have been forced for many years to comply with the government's strict regulation of producing cleaner, more fuel efficient vehicles. This research study focused on the implementation of the Corporate Average Fuel Economy standards (CAFÉ) and how that regulation impacted vehicles sales. Variables that were considered and processed through analysis are total vehicle sales post CAFE, fuel cost, and real GDP.

It was determined that automakers were forced to fund R&D as well as to develop new technology to meet the initial wave of standards. Most automakers took the associated cost of approximately $1000 in stride based on the competitive nature of car sales and the fact that they would lose market share if there was not compliance. Automakers found many ways to survive the regulation such as reducing the weight of vehicles, developing smaller, lower horsepower engines and passing additional the cost through to the customer. As the regulation evolved through the 1990's, Sport Utility Vehicles (SUV's) (which were not initially regulated) gained in popularity and production. They were quickly brought under regulations when their adoption by the public soared. The regulation opened new product lines such as the SUV and forced automakers to develop high performance, fuel efficient and clean burning alternatives to previous production.

Results from this study concluded that Government rules on the auto industry (fleet gas mileage **r**egulations on automakers) may have indirectly improved truck sales in conjunction with real GDP growth, while fuel prices directly impacted auto sales (but with an inverse relationship).

Background

In 1970, Congress passed the first major Clean Air Act, requiring a 90 percent reduction in emissions from new automobiles by 1975. Congress also established the U.S. Environmental Protection Agency (EPA), giving it broad responsibility for regulating motor vehicle pollution. New cars had to meet a 0.41 gram of hydrocarbons per mile standard and 3.4 grams of carbon monoxide per mile standard by 1975; nitrogen oxide emissions had to be reduced to 0.4 grams per mile by 1976. The nitrogen oxide standard was later revised in 1977. The Corporate Average Fuel Economy (CAFE) program established a phase-in of more stringent fuel economy standards beginning with 1975 model vehicles. Congress intentionally set technology-forcing standards, presenting automakers with major technical and economic challenges. The internal combustion engine was older technology that had not seen any substantial improvements in 20 years, and it was not clear that the standards could ever be met without replacing it all together. (Bresnahan 1985) Even if the necessary technologies emerged, the industry faced billions of dollars in R&D, capital and equipment, and installation costs. Congress and the EPA were aware of both the technological challenges and the high costs,

and no one was surprised by the contentious, adversarial nature of the implementation phase. Automakers incurred heavy costs of developing and adopting emissions controls, especially when the two new technologies were introduced in 1975 and 1981. The initial estimates for meeting the 90 percent reductions were $860 per vehicle. The fine levied on automakers was $10,000 per vehicle for non-compliance during a time when the average automobile sold for $2600 (Doyle, 1985).

The Corporate Average Fuel Economy (CAFE) regulations in the United States, first enacted by Congress in 1975, are federal regulations intended to improve the average fuel economy of cars and light trucks (trucks, vans and sport utility vehicles) sold in the U.S. in the wake of the 1973 Arab Oil Embargo. Historically, it is the sales-weighted average fuel economy of a manufacturer's fleet of current model year passenger cars or light trucks, with a gross vehicle weight rating (GVWR) of 8,500 pounds (3,856 kg) or less. The Energy Tax Act of 1978 in the U.S. established a gas guzzler tax on the sale of new model year vehicles whose fuel economy fails to meet certain statutory levels. The tax applies only to cars (not trucks) and is collected by the IRS. Its purpose is to discourage the production and purchase of fuel-inefficient vehicles. The tax was phased in over ten years with rates increasing over time. It applies only to manufacturers and importers of vehicles. Only new vehicles are subject to the tax, so no tax is imposed on used car sales. The tax applies a higher tax rate for less-fuel-efficient vehicles. To determine the tax rate, manufacturers test all the vehicles at their laboratories for fuel economy. The U.S. Environmental Protection Agency confirms a portion of those tests at an EPA lab.

The Act authorized the Secretary of Transportation to administer the CAFE program. The Secretary delegated the authority to the National Highway Traffic Safety Administration (NHTSA) Administrator. NHTSA was authorized to determine the maximum feasible CAFE levels; approve credit "carry back" and "carry forward" plans; determine and either grant or deny exemptions from the requirements for low-volume manufacturers; monitor the program through mandatory pre-model year and mid-model year manufacturer reports; and submit annually to Congress a report on the current status of the CAFE program. CAFE standards are to be declared 18 months prior to the beginning of the model year for which they are subscribed, with their determination established upon four basic statutory criteria: 1) technological feasibility; 2) economic practicability; 3) the effect of other Federal standards upon fuel economy; and 4) the need for the Nation to conserve energy.

The first year for which the standards were established for passenger cars was model year (MY) 1978 at a level of 18.0 miles per gallon (mpg); the standards increased to 19.0 mpg for MY 1979 and 20.0 mpg for MY 1980. The Act directed NHTSA to establish and declare standards administratively for MYs 1981, 1982, 1983 and 1984, and to specify fuel economy requirements for MYs 1985 and thereafter at 27.5 mpg. The fuel economy standard for light trucks was established for MY 1979 at 17.2 mpg for 2-wheel drive models and 15.6 mpg for models equipped with 4-wheel drive. Passenger vehicle calculations also changed over the years, and current calculations are made for manufacturers import fleets and the domestically produced fleet. The most recent fuel efficiency requirements for a manufacturer's passenger car fleets, either domestically produced or imported, is 27.5 mpg, the same performance level that was established by the Act for 1985**.** Light truck CAFE has been established through MY 2003 at 20.7 mpg. This level has remained unchanged since MY 1996 as a result of a rider incorporated into each year's Transportation and Related Agencies Appropriations Act that forbids NHTSA from changing the standard.

In 2007, the United States legislation expanded the Renewable Fuel Standard, mandating a significant increase in the use of bio-fuels by 2022 (Ford, 2010). In response, Ford Motor Company partnered with the U.S. Department of Energy, the Electric Power Research Institute, the New York State Energy Research and Development Authority and

Southern California Edison to explore expansion of electrification and other issues involved in expanding the use of plug-in hybrid electric vehicles. In 2009, the Department of Energy issued a conditional loan to Ford to finance up to 80% of qualified expenditures to improve efficiency of light vehicles by using technologies that improve internal combustion engines and transmissions, reduce vehicle weight, reduce vehicle drag with more aerodynamic designs, and improve vehicle efficiency through the development of hybrid and plug-in electric vehicles. The loan enabled Ford to raise the fuel efficiency of more than a dozen models, representing close to two million new vehicles annually, and saved more than 200 million gallons of gas a year. The DOE said the loan builds on steps taken by the Obama Administration to require an average fuel economy of 35.5 miles per gallon in the year 2016. The standard will reduce oil consumption by an estimated 1.8 billion barrels, prevent greenhouse gas emissions of approximately 950 million metric tons, and save consumers more than $3000 in fuel costs. The DOE extended conditional loan offers to Ford, Tesla Motors, and Nissan Motor for a total of $8 billion, in addition to offering $17 billion in loans to large and small auto manufacturers and parts suppliers up and down the production chain (Department of Energy, 2009).

Following the Energy Independence and Security Act of 2007 and administrative action by the Obama administration in 2009, standards will be aggressively tightened between 2011 and 2016. In fact, there are now two separate regulations: a National Traffic Highway Safety Administration (NHTSA) regulation that governs fuel economy, and an Environmental Protection Agency (EPA) regulation which limits CO2 emissions per mile. The regulations are essentially equivalent, except that the slightly tighter EPA requirement allows automakers to earn compliance credits by modifying air-conditioner refrigerants to reduce greenhouse gases. Automakers currently pay a fine of $55 per vehicle for every 1 mpg that their fleet average mpg falls short of the relevant standard, though presumably this fine will have to be increased to enforce the stricter standards.

The CAFE issue reaches well beyond Southeast Michigan in the United States. Every state in the union has jobs connected to automotive manufacturing. Whether with a manufacturer, supplier or the many small businesses that support these two industries, over 2.3 million workers across the nation rely on a healthy automotive sector. In reaction to this growing threat, the Chamber launched a national grassroots campaign to preserve the million of Automotive Jobs Action Coalition, known as AJAC, represents over 100,000 businesses around the nation. Over 30 chambers of commerce and business associations throughout the U.S. are now members of the coalition (U.S. Chamber of Commerce, 2008). Congress is poised to raise CAFE standards to a level that is not realistic, nor helpful, for the American automotive industry. The federal government should be creating incentives for consumers who purchase alternative fuel automobiles rather than passing regulations that are sure to push an industry teetering on collapse over the edge.

There is another way to improve fuel mileage and reduce the nation's dependency on foreign oil other than arbitrarily raising regulations to unattainable levels. Reasonable and responsible CAFE increases that provide automakers with the flexibility they need to meet higher standards without sacrificing cost, consumer choice and safety is a better route. In March 2011, the Detroit Free Press reported sluggish demand for fuel-efficient cars despite rising gas prices (Detroit Free Press, 2011). They published that hybrid car sales actually shrunk from 2.9 percent of new vehicle sales in 2009 to 2.4 percent last year and that sales of light trucks, SUV's crossovers and minivans, rose to 51% from 48% over the same period. They also reported that the during the first two months of the year, Chevrolet sold 602 Volts while Nissan sold 154 Leafs, while in contrast, Cadillac sold 2,793 Escalades and Lincoln sold 1,193 Navigators. Despite the government's fuel economy standard mandate that calls for a fleet-wide average of 35.5 mpg by 2016, the consumer demand for

large vehicles still remain strong. (See Appendix B).

Problem Statement– Hypotheses

H1car null: There is no correlation between Government Regulation of Auto Industry Fleet Gas Mileage Regulations (CAFÉ) vs. the sale of cars.
H1car alt There is a correlation between Government Regulation of Auto Industry Fleet Gas Mileage Regulations (CAFÉ) vs. the sale of cars.

H1truck null: There is no correlation between Government Regulation of Auto Industry Fleet Gas Mileage Regulations (CAFÉ) vs. the sale of light trucks.
H1truck alt: There is a correlation between Government Regulation of Auto Industry Fleet Gas Mileage Regulations (CAFÉ) vs. the sale of light trucks.

H2car null: There is no correlation between Real GDP vs. the sale of cars.
H2car alt: There is a correlation between Real GDP vs. the sale of cars.

H2truck null: There is no correlation between Real GDP vs. the sale of light trucks.
H2truck alt: There is a correlation between Real GDP vs. the sale of light trucks.

H3car null: There is no correlation between Average Annual Gas Prices vs. the sale of cars.
H3car alt: There is a correlation between Average Annual Gas Prices vs. the sale of cars.

H3truck null: There is no correlation between Average Annual Gas Prices vs. the sale of light trucks.
H3truck alt: There is a correlation between Average Annual Gas Prices vs. the sale of light trucks.

Research Design and Methodology

This assessment will use research and analysis to test the materiality of mandated automotive requirements and industry data. To understand industry response to regulations, historical data was examined including CAFE standards, car unit sales, light truck unit sales, Real GDP, and annual average gas price. Using vehicle unit sales of cars and light trucks as the dependent variables, a backward stepwise regression analyses was performed to determine if a significant relationship exists vs. the various independent variables, how strong and in what direction these relationships are, and what regression model could be generated to predict vehicle sales.

Government intervention, with industries such as automotive, is assumed necessary due to the government's initiatives to drive down the dependence of foreign oil, improve climate change, focus on energy sustainability, and save consumers money. However, by government's reaching hand to create legislation that dictates the industry standard, the industry is forced to either invest capital to meet the requirements, risk heavy fines, lose a competitive edge, and/or pass along costs to the consumer.

The data used for this analysis was sourced from the following:

- National Highway Traffic and Safety Administration, 2010.
- Ward's Auto, a leader in providing automobile data for more than 80 years Ward's Auto, 2009.
- US Department of Commerce, Bureau of Economic Analysis, 2011.
- US Department of Energy, 2010.

The National Highway Traffic and Safety Administration provided key information relating to CAFE standards over the course of 27 years (1978 – 2004). For this analysis, it was important to analyze data prior to implementation of CAFE standards in 1978. Additionally, we extrapolated data from Ward's Auto for car unit sales, US Department of Commerce for Real GDP and US Department of Energy for historic annual average gas prices.

Statistical tests were performed to test the strength of their relationships.

Multiple Regression Analysis for U.S. Car Sales

Table 1: Model Summary

Multiple R	0.608
R Square	0.370
Adjusted R Sq	0.344
Standard Error	963.324
Observations	27

Predictors: AvgAnnualGasPrices

Table 2: ANOVA Statistics

	SS	df	MS	F	Sig
Reg.	1.360E7	1	1.36E7	14.653	.001
Res.	2.320E7	25	927993		
Tot.	3.680E7	26			

Predictors: AvgAnnualGasPrices
Dependent Variable: CarSalesPerThousand

Table 3: Coefficients

	B	t	Sig.
Constant	12858.78	12.829	.000
GasPrice	-31.95	-3.828	.001

Dependent Variable: CarSalesPerThousand

While CAFÉ and Real GDP were not found to be a significant predictor of car sales, Average Annual Gas Prices was. The variable had an inverse relationship to the dependent variable, so an increase in average annual gas prices corresponded to a decrease in car sales. The predictive model for car sales is:

*Car sales = 12858.78 – 31.95 * Avg Gas Prices*

Multiple Regression Analysis for U.S. Light Truck Sales

Table 4: Model Summary

Multiple R	0.980
R Square	0.961
Adjusted R Sq	0.958
Standard Error	452.458
Observations	27

Predictors: CAFE, RealGDPinBillions

Table 5: ANOVA Statistics

	SS	df	MS	F	Sig
Reg.	1.213E8	2	6.067E7	296.4	.000
Res.	4.913E6	24	2.047E5		
Tot.	1.263E8	26			

Predictors: CAFE, RealGDPinBillions
Dependent Variable: LtTruckSalesPerThousand

Table 6: Coefficients

	B	t	Sig.
Constant	-7398.99	-8.653	.000
RealGDP	.85	15.127	.000
CAFÉ	187.87	4.505	.000

Dependent Variable: LtTruckSalesPerThousand

While Gas Prices were not found to be a significant predictor of light truck sales, both Real GDP and CAFE were. Both variables had a direct relationship to the dependent variable, so an increase in Real GDP corresponded with an increase in light truck sales, and likewise an increase in CAFE also corresponded to an increase in light truck sales. The predictive model for light truck sales is:

*LtTruck sales = -7398.99 + 0.848 * RealGDP + 187.87 * CAFE*

Results and Conclusions

Appendix C graphically illustrates the sales of cars and light trucks during the test period. The results of the hypothesis tests showed fuel prices

to be a significant predictor of car sales in the U.S., with an increase in fuel prices corresponding to a decrease in car sales. The correlation of .608 indicates that fuel prices is a moderately strong predictor, and may in fact be a direct cause of the increase or decrease in car sales. With virtually every car running on gasoline, it makes sense that car sales depend heavily on gas prices. The dramatic increase in gas prices over the past decade has hurt the sales of many cars, but hybrid cars and diesel vehicles have fared much better.

Unlike the cars, light truck sales in the U.S. were not significantly affected by gasoline prices, but were positively related to the real GDP and the CAFÉ standards. Possible explanations of the truck sales' relationship with these three factors follow. As most trucks have diesel engines and diesel vehicles get about 30% better fuel economy, while diesel prices are about 10% higher than gasoline, it makes sense that gas prices would not be a factor in light truck sales. With CAFÉ standards being raised, hybrid cars and diesel vehicles are readily meeting requirements and so an increase in the sale of light trucks makes logical sense. Real GDP growth in the last ten years of the data coincided with the phenomenal growth of SUVs. SUVs have become mainstream and may have acted as a catalyst in engendering more truck sales. In the past, trucks were often considered farm / ranch vehicles; today, trucks are commonly seen in cities and suburban areas. Thus a social change in the acceptance of trucks could have been a factor in stimulating truck sales at the same time that real GDP was increasing.

Table 7: Summary of Research Results

	Significant factors	
	Cars	Light Trucks
CAFE Regulations		✓
Real GDP		✓
Cost of Fuel	✓	

Recommendations

Based on the results of the research performed, CAFÉ regulations, real GDP and fuel costs are all significant in the analysis of car sales or light truck sales. While light truck sales were related positively to two of the variables, car sales were related inversely to the third. It seems that the major difference is that cars were predominantly gasoline engines while light trucks were predominantly diesel. In the last few years the sale of hybrid cars has grown significantly while gasoline prices and the economy have worsened. As such, it is recommended that further research be conducted to include the most recent decade of car and truck sales, and it should break down vehicles as gasoline-powered, diesel and hybrid, and possibly new vs. used vehicles.

Other factors that could be analyzed in more expanded future studies include quality, customer satisfaction, price, financing offers, and buyer preferences. Clearly in bad economies, there is a concern for getting the best fuel economy and getting value for the purchase. Auto dealers can certainly benefit greatly from knowing what vehicles to market and carry to meet likely demand.

REFERENCES

[1] Bresnahan, T. F. and Dennis A. Y. (1985). "The no pecuniary costs of automobile emissions standards," Rand Journal of Economics, 16(4):437-455.

[2] Carolyn, A. (2011). "Sluggish demand for fuel efficient cars despite rising gas prices". Detroit Free Press.

[3] Doyle, J. (2000). Taken for a Ride: Detroit's Big Three and the Politics of Pollution. New York: Four Walls Eight Windows.

[4] Ford Motor. (2010). Sustainability *report 2009/10*. Retrieved from Ford Motor.com on Mar 3, 2011 http://corporate.ford.com/microsites/sustainability-report-2009-10/issues-climate-performance-fuel

[5] National Highway and Traffic Safety

Association, (2010). http://www.nhtsa.gov/fuel-economy

[6] North, D. C. (1990). *Institutional Change and Economic Performance*. Cambridge University Press, 1990.

[7] United States Department of Energy. (2009). *US Energy Secretary Chu Announces Finalized $5.9 Billion Loan for Ford Motor Company.* Retrieved from the United States Department of Energy on March 31, 2011. http://www.energy.gov/news/8023.htm

[8] US Chamber of Commerce, (2008). Automotive Jobs Action Coalition. http://www.usgovernmentspending.com

[9] US Department of Commerce, Bureau of Economic Analysis, (2011). http://www.usgovernmentspending.com/us real gdp history

[10] US Department of Energy, (2010). http://www.epa.gov/oms/inventory/overview/solutions/milestones.htm

[11] Ward's Auto, (2009). http://www.wardsauto.com/keydata/historical/UsaSa01summary.xls

[12] Whoriskey, P. (2010). *Light bulb factory closes; end of era for US means more jobs overseas.* The Washington Post. Sept 8, 2010.

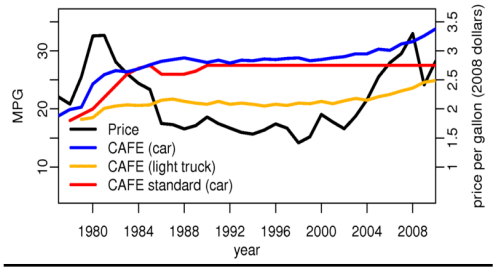
APPENDICES

Appendix A – Auto Sales Data

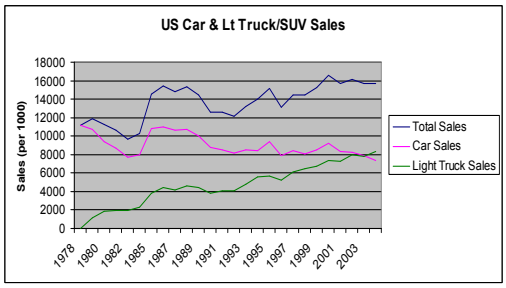
Year	Car Sales*	Truck Sales*	CAFE	Real GDP**	Avg Gas
1978	11182	0	18.0	5678	63
1979	10747	1165	19.0	5855	86
1980	9390	1885	20.0	5839	125
1981	8687	1950	22.0	5987	138
1982	7752	1948	24.0	5871	130
1983	7967	2309	26.0	6136	124
1984	10700	3445	27.0	6577	121
1985	10796	3779	27.5	6849	120
1986	11017	4423	26.0	7087	93
1987	10671	4166	26.0	7313	95
1988	10695	4604	26.0	7614	108
1989	10005	4448	26.5	7886	112
1990	8815	3802	27.5	8034	129
1991	8530	4058	27.5	8015	110
1992	8143	4048	27.5	8287	109
1993	8480	4760	27.5	8523	107
1994	8423	5565	27.5	8871	107
1995	9434	5711	27.5	9094	110
1996	7920	5224	27.5	9434	119
1997	8384	6098	27.5	9854	119
1998	8031	6453	27.5	10284	102
1999	8517	6718	27.5	10780	112
2000	9247	7333	27.5	11226	146
2001	8347	7315	27.5	11347	138
2002	8236	7945	27.5	11553	131
2003	7894	7822	27.5	11841	152
2004	7367	8346	27.5	12264	181

***per thousand **in billions**

Appendix B – Price of gas/miles per gallon



Appendix C – Car Sales



A Low Cost Solar Tracker Design for Renewable Energy

Jeng-Nan Juang, R. Radharamanan, and Jason Beaver
Mercer University
juang_jn@mercer.edu, radharaman_r@mercer.edu

Abstract

A solar tracker is a device that orients payloads such as photovoltaic (PV) panels, reflectors, lenses or other optical devices toward the Sun. In flat-panel PV applications, trackers are used to minimize the angle of incidence between the incoming sun light and a PV panel to increase the amount of energy produced from an installed fixed capacity power generating units. The sunlight has two different components: the "direct beam" component and the "diffuse sunlight" component. The "direct beam" carries about 90% of the solar energy and the "diffused beam" carries the remainder. Since the majority of the energy is in the "direct beam", to maximize the energy collection the Sun should be visible to the solar panels as long as possible.

Since 2009, in standard PV applications, trackers are used in at least 85% of commercial installations producing greater than 1MW. Trackers are used to enable the optical components in both concentrated photovoltaic (CPV) and concentrated solar thermal (CSP) applications. The optics in CSP applications accept the direct component of sunlight and therefore must be oriented appropriately toward the Sun to collect energy. All concentrated applications require tracking systems because such systems do not produce energy unless oriented toward the Sun.

In this study, a low cost solar tracker for various payloads has been designed for flat-panel and concentrated PV applications in order to collect maximum energy from "direct beam" as well as "diffuse sunlight". The design details are presented and discussed.

1. Introduction

The angle of beam radiation is constantly changing since the Sun changes daily, monthly, and seasonally the way it carves its path across the sky. The solar panels should be oriented perpendicular to the Sun at all times to maximize output from the solar panels. A solar power tracing system helps to achieve this [1].

The amount of energy gained will depend on the location and the type of tracking system used. One can see from the solar energy tables for the state of Washington, the difference between a fixed solar panel system and one with tracking was approximately 40% more in the summer months, a 25% increase in the spring months, and about 10% increase in the winter months [2].

A manually adjustable one axis system is the simplest type of tracking. In this system, the solar panels are installed in a south facing direction, with a rotating mechanism on their horizontal axis. Making seasonal adjustments to maintain the best operating angle for that time of year, this simple tracking system can gain 20% better performance when compared to a fixed solar panel installation [2].

Automatic (active) one axis tracking systems are self-adjusting units that rotate daily with the arc of the Sun. This system can provide performance gains of 40% or more in the summer months. This system has the added complexity of the electronic and mechanical control mechanisms. The additional cost of this system can be offset by their superior performance [3].

Two axis tracking systems are the most complex and most accurate installations available. In this system, the solar panels have additional flexibility on their vertical axis (East to West). These systems are the most exact when it comes to following the Sun and is used with systems requiring high temperatures on the receiving unit.

A good tracking is needed in getting the most out of the solar panels. One axis systems are effective, inexpensive, and easy to use. Two axis systems are more complex and normally used in high temperature systems [3].

The drive types used in the tracking systems include the following: 1. Active trackers use motors and gear trains to direct the tracker as commanded by a controller responding to the solar direction; 2. Passive trackers for photovoltaic solar panels use a hologram behind stripes of photovoltaic cells so that sunlight passes through the transparent part of the module and reflects on the hologram. This allows sunlight to hit the cell from behind, thereby increasing the module's efficiency; 3. Chronological trackers counteract the Earth's rotation by turning at an equal rate as the earth, but in the opposite direction; and 4. Manual tracking - in some developing nations, drives have been replaced by operators who adjust the trackers. This has the benefits of robustness, having staff available for maintenance, and creating employment for the population in the vicinity of the site [4].

When developing a solar tracker there are two main orientations that need to be addressed, azimuth and the angle of inclination. The azimuth of an object is its direction in the sky, measured in degrees. It corresponds to the cardinal direction on land, namely north (0 or 360 degrees), east (90 degrees), south (180 degrees) and west (270 degrees). It is normally defined as the angular distance along the horizon between a point of reference, usually the observer's bearing, and another object. In this case the other object is the Sun. When thinking about the Sun, one can think of azimuth as the east to west path that the Sun appears to travel upon from sunrise to sunset. Azimuth calculation is critical since the Sun will never stay in one place for long (Fig. 1).

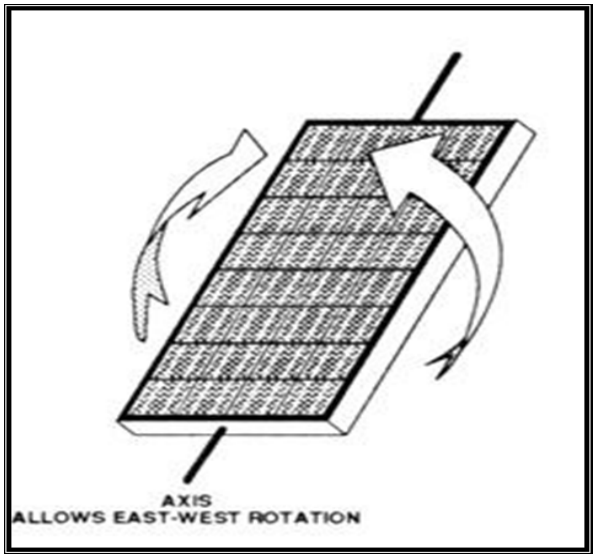


Figure 1. Azimuth

On the other hand, the angle of inclination (Fig. 2) is just as important when it comes to tracking the Sun. Although the Sun follows a path that can be calculated fairly easily, the path is constantly changing from month to month. As one can see in the Fig. 2, a stationary solar collector will gather solar power in varying degrees of efficiency, whereas a tracker has the ability to adjust its angle in order to minimize the angle of incidence between the incoming light and the panel itself [4-6].

2. East-West orientation

The first phase of the design process deals with the azimuth aspect of the solar panel mounting system. The East-West Orientation device (Fig. 3) needs a good range of motion from left to right so that the panel will be able to follow the path of the Sun regardless of its initial orientation. The current plan is to go with an antenna rotator, which is a device that is normally used to change the orientation of a directional antenna. This device will give the tracker the full range of motion that it needs and the cost of the device is fairly low. Also, the device is programmable and since it is made for outdoor use, it is ideal for this project [5].

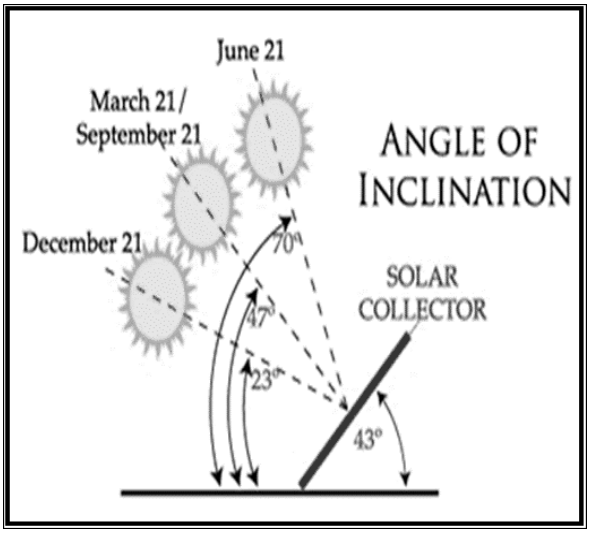


Figure 2. The angle of inclination

3. North-South orientation

The second phase of the design process deals with the angle of inclination aspect of the solar panel mounting system. A North-South Orientation device (Fig. 4) is needed that will tilt the panel depending on the position of the Sun as it progresses in the sky throughout the day. An easy solution to this problem can be found in the implementation of a programmable linear actuator [4]. By attaching the actuator's telescoping rod to a key position on the panel it will be able to achieve the desired tilt. This particular model of linear actuator is waterproof and therefore, it is ideal for this project.

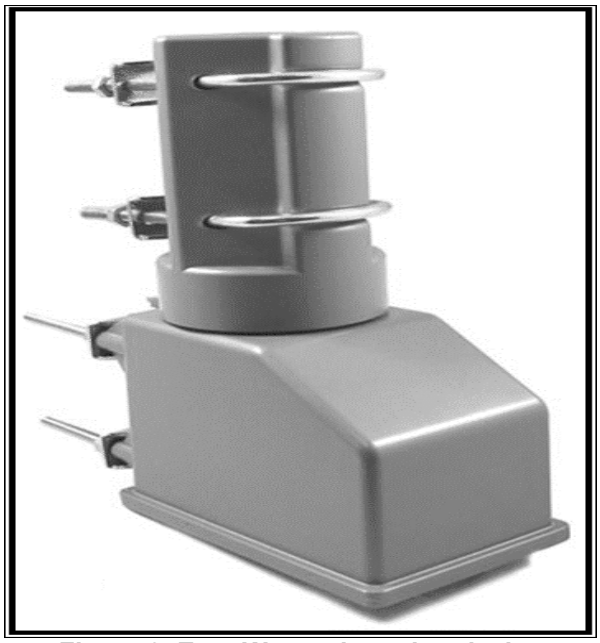


Figure 3. East-West orientation device

4. Light dependent resistors

The third phase of the design requires one to devise an efficient way to determine whether or not the antenna rotator and the linear actuator should even activate in the first place. The reason one would want to do this is because the device should only operate in the event that there is a sufficient amount of light available for the panel to absorb. For instance, if it is extremely cloudy outside then there is no real reason for the panel to move. Moreover, the tracker needs a way to measure the available amount of light. The solution to this problem will be found through the use of photo resistors or light dependent resistors (LDR). An LDR is a variable resistor that changes its resistance based on the amount of ambient

light that falls onto its surface (Fig. 5). By using the LDRs to confirm that there is a significant amount of ambient light available, the system will know whether or not it needs to activate [7-9].

Figure 4. North-South orientation device

5. Raspberry Pi

The fourth phase of the design project will deal with the incorporation of a small onboard computer into the solar panel mounting system. The system that will be used is a credit card-sized single-board computer called Raspberry Pi (Fig. 6). This tiny computer uses a Linux based operating system and it will essentially be the brains of the solar tracker. Initially, the Pi will accept inputs from the light dependent resistors. If the inputs suggest that a sufficient amount of light is present, then the Pi will access a computer program called Stellarium. By accessing the program the Pi will be able to communicate with the antenna rotator and the linear actuator [10].

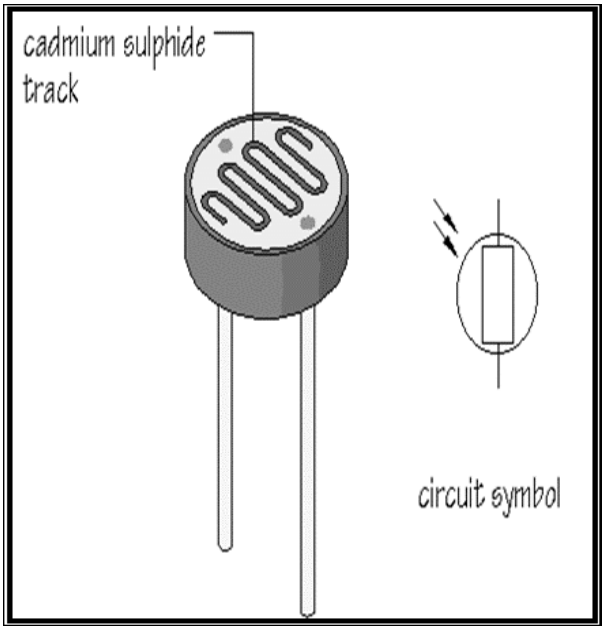


Figure 5. Light dependent resistor

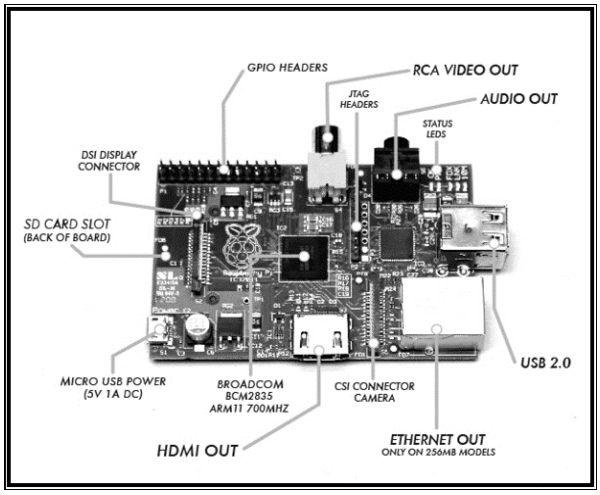


Figure 6. Raspberry Pi board

6. Stellarium

Stellarium is a free software planetarium program that is compatible with Linux based operating systems [10]. The main usage of this program deals with a real-time projection of the night sky. However, the program also has a very realistic depiction of the Sun from sunrise to sunset. By entering in the latitudinal and longitudinal coordinates, one can see exactly

where the Sun will be at anytime during the day. Therefore, the tracker will have a virtual knowledge of the azimuth and altitude of the Sun at any theoretical moment in time. Since the tracker will have an exact location to go to, it will not need to waste time looking for the Sun (Fig. 7). Instead, the tracker will orient itself towards the point where the Sun should be. This will reduce any unnecessary motion that the tracker might go through if it were relying solely on the photo resistors.

Figure 7. Tracking sunset

7. Potential Design

The solar tracker system consists of: base, tubing, raspberry Pi and Stellarium software planetarium program, antenna rotator, linear actuator, light dependent resistors, and solar panels [11-14]. The potential design of the system is shown in Fig. 8.

The tracker system will function as follows: During the day, if sufficient sunlight is detected, the LDRs will activate the antenna rotator and the linear actuator. Then the Raspberry Pi will access the Stellarium computer program to send altitude coordinates to linear actuator and Azimuth coordinates to antenna rotator to track the location of the Sun and position the solar panels appropriately to collect maximum energy from solar radiation. The cycle will be repeated every 10 minutes when sufficient sunlight is detected by the LDRs. If sufficient sunlight is not detected by the LDRs, the system will stop and no energy will be collected until sufficient sunlight is available. This process will continue during the day and for consecutive days. The operations flow chart is shown in Fig. 9.

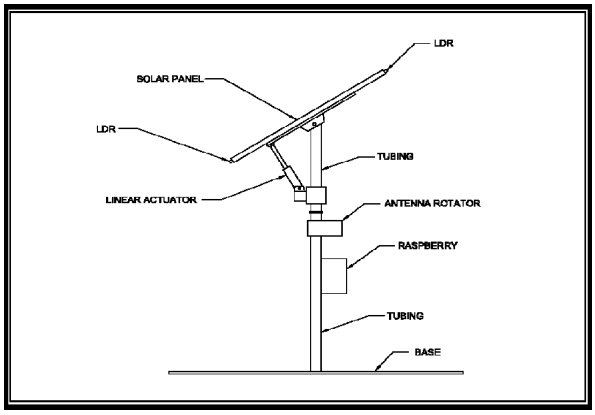


Figure 8. Potential design sketch

8. Conclusions and future work

A low cost solar tracker for various payloads was designed for flat-panel and concentrated photovoltaic applications to collect maximum energy from “direct beam” and “diffuse sunlight” by varying the angle of incidence between the incoming light and a photovoltaic panel.

Once the tracker has been built and tested, it will be positioned on top of Mercer University’s School of Engineering building. Once there, the device will be alternated between tracking and non-tracking modes in order to determine the efficiency of the proposed device. In addition, a digital log of the trackers movements will be kept in order to determine whether or not it is functioning properly after consecutive days in the field. Ultimately, this solar tracker will be

integrated into a larger system for a senior design project.

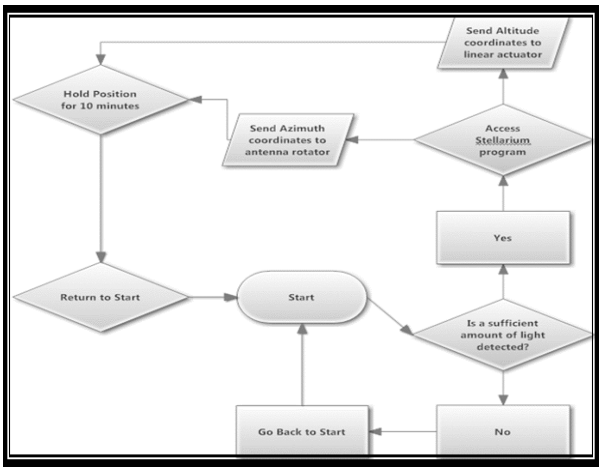


Figure 9. Operations flowchart

The future work is to promote the use of renewable sources of energy by designing and developing a semi-portable Real Time Solar Energy System (RISES). This product will provide an alternative source of power to a standard sized home and also serve as a stand-alone source of power in remote areas. This system will be designed to power electrical gadgets such as a laptop, a television, and a lamp. This product will serve as an advocate for green energy and will help reduce the pressure on electricity generated by fossil fuels and nuclear energy by providing a renewable alternative [15]. The goal is to increase the use of solar energy as a source of electricity in various regions. The target customers include but not limited to: people who desire an alternative source of electricity in their homes; people in remote areas who need a source of power; and people who are advocates of green energy.

9. References

[1] Solar Power Tracker: Renewable Energy Goods, Expertise and Technical Advice, altEstore.com, http://www.altestore.com, Retrieved March 2, 2013.

[2] W. D. Lubitz, "Effect of Manual Tilt Adjustments on Incident Irradiance on Fixed and Tracking Solar Panels", Applied Energy, Volume 88 (2011), pp. 1710-1719.

[3] D. Cooke, "Single vs. Dual Axis Solar Tracking", Alternate Energy eMagazine, April 2011.

[4] Solar Actuators: Manufacturer of Solar Tracking Actuators for PV, CPV, & CSP, joycedayton.com, http://www.joycedayton.com, Retrieved March 5, 2013.

[5] Programmable Antennas: http://www.antennasdirect.com/store/one_cable_programmable_antenna_rotator.html, Retrieved February 20, 2013.

[6] http://www.northerntool.com, Retrieved March 5, 2013.

[7] http://www.Amazon.com/Photoresistor, Retrieved March 18, 2013.

[8] Andreev, V. M., V. A Grilikhes, and V. D. Rumiavtsev, Photovoltaic Conversion of Concentrated Sunlight, John Wiley, 2007.

[9] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Analysis of Factors Influencing the Annual Energy Production of Photovoltaic Systems", Photovoltaic Specialists Conference, May 2002, Conference Record of the Twenty-Ninth IEEE: 1356–1361.

[10] http://stellarium.org, Retrieved February 3, 2013.

[11] Goswami, D. Y., F. Kreith, and J. F. Kreider, Principles of Solar Engineering, Taylor & Francis, Philadelphia, PA, 2000.

[12] Masters G. M., Renewable and Efficient Electric Power Systems, 4th edition, John Wiley, 2004.

[13] Duffie J. A., and W. A. Beckman, Solar Engineering of Thermal Processes, John Wiley, 2006.

[14] Thomas I., The Pros and Cons of Solar Power, Rosen Central, 2007.

[15] Kothari, D. P., and I. J. Nagrath, Modern Power System Analysis (3rd edition), Tata McGraw-Hill Publishing company, New Delhi, 2003.

iFTDM: A New Fault-Tolerant Routing Algorithm Utilizing Unicast-based and Tree-Based Multicasting in Mesh Including Number of Fault Regions and Fault Region Size

Masoud E. Shaheen[1], Ahmed Abukmail[2] and Dia Ali[3]
[1]Fayoum University, [2]University of Houston - Clear Lake, [3]University of Southern Mississippi
Masoud.shaheen@eagles.usm.edu, Abukmail@uhcl.edu, dia.ali@usm.edu

Abstract

Multicast communication services have numerous uses in distributed memory systems, large-scale multiprocessors and parallel computer systems. One of the fundamental problems in parallel computing is how to efficiently perform routing in a faulty network each component of which fails with some probability. In previous research a fault tolerant deadlock-free multicast routing algorithm for wormhole routed 2D mesh multicomputer without considering the overlapping of fault regions was proposed and without considering the effect of the number of fault regions and the fault region size [1]. In this paper, an improved fault tolerant deadlock-free multicast routing algorithm, for wormhole routed 2D mesh multicomputer, *i*FTDM, is presented. The number of fault regions and the fault region size affects are considered. A simulation study is conducted to measure the total number of network traffic steps and network latency steps.

1. Introduction

In a multicomputer system, a collection of nodes (or processors) participate with each other to resolve difficult application problems. These nodes communicate their data and synchronize their efforts by sending and receiving packets through the essential communication network. One of the important issues in parallel computing (distributed memory systems) is how to powerfully accomplish routing in a faulty network, where each element fails with various probabilities. Routing is a job where a source node sends a message to destination nodes. Network topology is an important factor that affects routing algorithms.

Mesh networks have been extensively used in most distributed memory systems. Mesh networks are chosen because they offer advantageous edge connectivity, which can be used for input/output controllers [2]. The main idea of wormhole routing is that a packet is divided into flits. That makes a packet (message) delivered easily to destination nodes. Fault tolerance routing techniques are an essential issue facing the design and implementation of interconnection networks for distributed-memory systems. An ideal message fault tolerant multicast routing should achieve minimum traffic steps for the interconnection networks involved. Unfortunately, finding optimal message routing has been shown to be NP-hard for most common distributed memory systems topologies. The occurrence of fault regions renders existing routing algorithms to deadlock-free routing useless. This paper will

focus on studying the fault-tolerant multicast wormhole routings in 2D mesh networks with overlap on convex fault regions.

The rest of this paper is organized as follows: In section 2, some related work is shown. Section 3 gives background of the work. Section 4 presents the proposed fault-tolerant routing method. Section 5 shows results and performance analysis. Finally in Section 6, we will give some conclusions for the results obtained in this paper.

2. Related work

Fault tolerant routing in direct networks has been gaining attention in recent years. The model of individual link and node failures produces patterns of failed elements. Fault regions result from the closest faulty links and faulty nodes. The three most important faults are convex, concave, and irregular. A good fault tolerant routing should be simple (low implementation cost) and deadlock-free. Also, the effect of number of fault regions, number of destinations and fault region size needs to be taken into consideration. Furthermore, all these goals should be achieved with less consideration for hardware requirement.

Fukushima et al [3] proposed a Hardware-Oriented fault tolerant routing algorithm for irregular 2D mesh network. The proposed algorithm requires no virtual channels per physical channel to ensure deadlock-free in irregular 2D mesh. The proposed Position-Route algorithm for non-VC routers requires much less routing complexity. The basic idea is to integrate routing behaviors of the traditional message-based algorithm and to simplify the ring selection, but the effect of number of fault regions is not presented. Dai et al [4] proposed, MORT, a technique to develop routing efficiency in fault tolerant multipath routing. MORT is based on a technique called information hiding. This technique permits routers in network to hide some routing information, such as link failures and link cost changes to other routers deprived of a bad effect to the routing protocols. MORT reduces the message overhead and convergence time in the multipath routing protocol design by calming the limitations on the shape of the faulty regions and the size of fault region.

Rezazadeh et al [5] proposed a performance improving fault-tolerant routing algorithm for Network-on-Chip in uniform traffic based on f-cube3 as a solution for the accumulative rate of switched and routed packets in Network-on-Chip. It is proposed that when a message is not blocked by fault, all virtual channels could be used. The proposed algorithm requires only one virtual channel per physical channel to ensure deadlock-free. Also, Rezazadeh et al [6] proposed an improved fault-tolerant routing algorithm for mesh Network-on-Chip. The proposed algorithm requires only two virtual channels per physical channel to ensure deadlock-free. The suggested modification tolerates multiple block faults with overlapped f-rings. An entire column/row fault disconnects meshes and size of fault region is not considered.

Mejia et al [7] proposed an efficient fault tolerant routing algorithm for meshes and tori networks. The proposed algorithm is a deterministic routing methodology for tori and meshes, which achieves high performance without needing virtual channels. This algorithm can deal with any topology derived from any combination of faults when combined with static reconfiguration. The algorithm, called Segment-based Routing, which works by partitioning a topology into subnets, and subnets into segments, size of fault is not considered. In [8] Chang and Chiu presented fault tolerant unicast-based multicast routing algorithm, FT-cube2, in

2D meshes. In the FT-cube2, e-cube routing algorithm is enhanced in order to deal with multiple fault regions in 2D meshes using only two virtual channels per physical channel. In FT-cube2 routing algorithm, normal messages are routed through e-cube hops. A message is delivered on an f-ring or f-chain to destination nodes along clockwise or counterclockwise directions. FT-cube2 is the compared routing algorithm with this work.

3. Background

This section first give brief review to multicast routing algorithms, then an introduction to column path routing.

3.1. Multicast routing algorithms

There are three basic types of multicast routing algorithms: unicast-based, tree-based and path-based [9], but in this paper a new type is constructed: *i*FTDM (*i*mproved FTDM), which is a compromise of tree-based and unicast-based, called unicast/tree based routing. In unicast based algorithms, a source node routes a message to the destinations by sending a series of separate unicast messages to each destination. It needs a startup for each destination. The separate addressing is a unicast based multicast routing, in which the source node sends directly a separate copy of the message to every destination node [10].

Multicast tree-based routings deliver a message from source node to destination nodes in a single multi-head worm that splits at some routers and replicates the data on multiple output ports. Multicast path-based routing permits a worm to hold a sorted list of several destination node addresses in its header flits. They use a simple hardware mechanism to allocate routers to absorb flits on interior channels while they concurrently forward copies of the flits on output channels sending to the remaining destination nodes.

3.2. Column path routing

In [11], Boppana et al. proposed a multicast routing algorithm called the column path algorithm for mesh and torus networks. In a k x k 2D mesh, the column path algorithm partitions the set of destinations of a multicast message into at most 2k, such that there are at most two message copies directed to each column in the mesh. One for destinations on high levels (rows) above the source node, and the second copy for destinations on low levels (rows) below the source node. If a column holds one or more destinations of a multicast communication in the same row or in rows above that of the source, subsequently one message copy will be sent to service all those destinations. Likewise, if a column holds one or more destinations of a multicast communication in the same row or in rows below that of the source, then one message copy will be sent to service all those destinations. If all destinations of a column are either below or above the source node, then one message copy will be sent to service all those destinations. Messages are routed using the row-column or e-cube routing method, therefore, the column path routing algorithm is well-matched with the e-cube algorithm. Figure 1 illustrates an example of the column-path algorithm on a 2D mesh. We consider node (4, 4) as a source node.

The column path routing algorithm has proved a deadlock free routing algorithm [11] because the source node sends directly a separate copy of the message to every destination node on the same column, then no cyclic dependency can be created among the channels.

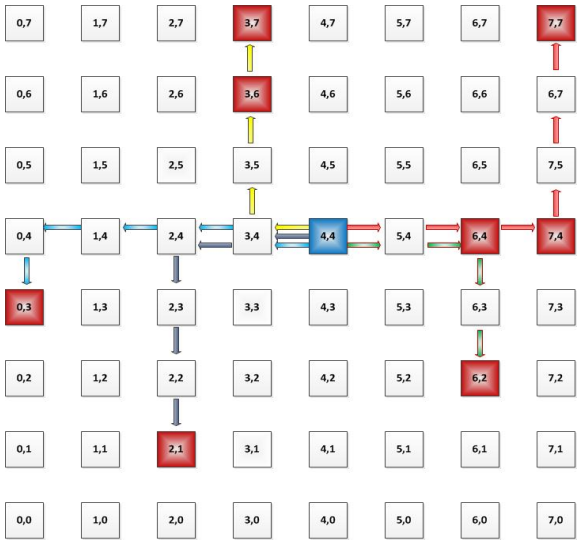


Figure 1. An example of the column path routing on a 2D mesh

4. Improved algorithm (iFTDM)

Most fault tolerant routing algorithms which were proposed recently concentrate on unicast-based multicast algorithms. Unicast-based algorithms require a startup time for each destination, and this requires more work. Also, they are incompetent because they permit a message to be delivered to only one destination, which leads to multicast operations being implemented as multiple phases of multicast message exchange. So, contention freedom must be guaranteed not only among the worms of a given phase, but also among worms in different phases [1].

Furthermore, many fault tolerant routing algorithms which were proposed recently concentrate on the number of destinations as the main parameter that must be considered calculating network latency steps and network traffic steps in mesh networks. On the other hand, the effect of number of fault regions and fault region size are not taken into consideration when they calculate network latency steps and network traffic steps for their algorithms.

In this section, a new fault tolerant deadlock free multicast routing algorithm, *i*FTDM, for 2D meshes is introduced. *i*FTDM is a unicast/tree-based multicast algorithm, which attempts to deliver the message to all destinations in two phases, the same method as FTDM work. In the first phase, the message is delivered as multicast unicast-based to X-coordinate nodes (nodes (0, y_{bi}) in case of odd rows or (m-1, y_{bi}) in case of even rows) of each *true fault regions* at these nodes; *central nodes*. Each node is considered as a source node that has a message with header containing destinations in the three locations around the fault. In the second phase, the message is delivered from the central nodes in multicast tree-based fashion, which attempts to route the message to all destinations in a single multi-head worm that splits at some routers and replicates the data on multiple output ports. In a 2D mesh L_1, L_2 and L_3 are three locations around each true fault region as in figure 2, and L_4 is a location in case the 1st fault region is an f-ring.

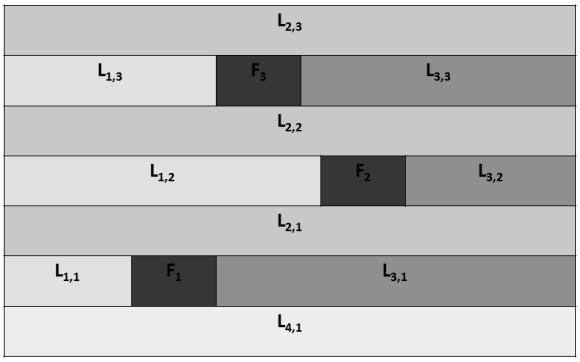


Figure 2. Locations around fault regions

4.1. Overlapping convex fault regions

The proposed fault tolerant multicast routing algorithm, *i*FTDM, does tolerate overlapping convex faulty regions. If there is an overlapping region, then no L_2 will form in between the overlap faults, at this point there are two cases of overlapping, the 1st case is overlap on L_1 which y_b of F_{i+1} less than y_e of F_i as shown in figure 3(a), the 2nd case is overlap on L_3 which y_b of F_i

less than y_e of F_{i+1} as shown in figure 3(b). In figure 3, star shapes represent central nodes.

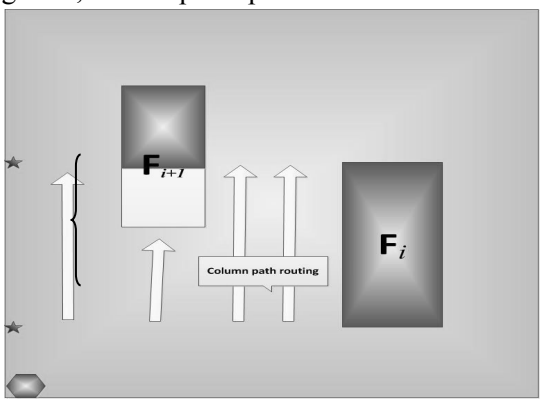


Figure 3(a). Overlapping on L_1

In order to tolerate the overlapping, the column path routing technique described before on background section needs to be used.

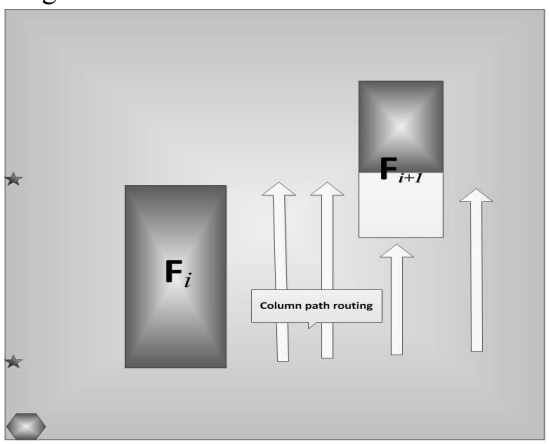


Figure 3(b). Overlapping on L_3

4.2. Routing functions

The *i*FTDM routing algorithm uses the same routing functions, R and R FTDM used, both of which proved deadlock-free. *i*FTDM assigns a label for each node based on the position of that node in a Hamiltonian path [12] as follows:

$$Q(p_i) = Q(x_i, y_i) = \begin{cases} y_i \quad n + x_i \quad 1 & y_i \text{ is even} \\ y_i \quad n + n - x_i & y_i \text{ is odd} \end{cases}$$

R (c, d) = w, where

$$Q(w) = \begin{cases} \max\{Q(z): Q(z) \quad Q(u)\} \text{ if } Q(c) < Q(d) \\ \max\{Q(z): Q(z) \quad Q(u)\} \text{ if } Q(c) > Q(d) \end{cases}$$

and z is a neighboring node of c

R (c, d) = w, where

$$w = \begin{cases} (x_c, y_c - 1) & \text{if } x_d = x_c \\ (x_c, y_c + 1) & \text{if } x_d = x_c + d_x \quad d_{Fi} \\ (x_c + d_x, y_c) & \text{otherwise} \end{cases}$$

Lemma*: i*FTDMalgorithm is deadlock-free

Proof: Because *i*FTDM uses R, R ,and column path routing to route a message on a 2D mesh with convex faults and all of these are deadlock-free (no cyclic dependency can be created among the channels) as mentioned. Then *i*FTDM is deadlock-free.

Algorithm (FTDM vs iFTDM)

In previous research [1] we introduced FTDM algorithm,

Input: The message *mess***,** Label node LN, central nodes CN_k, destination set D, and fault region F_i .

Output: d_j D, Receive(d_j, mess)

Procedure:

[1]/* Phase 1 (unicast-based): Send copies of message to CN_k

[a] If c = d_1 then

a. 1) D = D-{c}

a. 2) Receive(c, *mess*)

[b] If D = then stop

[c] Send separate messages to CN_k using XY routing.

[d] Modify header of messages, *mess*, and put in each header D_k destinations, which k is the number of central nodes (plus one if first fault is f-ring)

[e] Let each CN_k as a new source node

[f] Go to phase 2

[2]/* Phase 2 (tree-based):

[a] If $c = d_1$ then
a. 1) D = D-{c}
a. 2) Receive(c, *mess*)
[b] If D = then stop
[c] At each new source node, send two copies of message, *mess,* we have three cases:
Case 1 (No overlap)
c. 1) 1st copy contains destinations on L_1 and L_2 using R
c. 2) 2nd copy contains destinations on L_3. Using R' to route a message around the fault region until the message reach to LN, and then use R.
c. 3) If L_3 have another faults then recursively apply *i*FTDM.

In this research we proposed the following improvement (iFTDM) algorithm:
Case 2 (overlap in L_1)
c. 1) 1st copy contains destinations on L_1 using *column path* routing.
c. 2) 2nd copy contains destinations on L_3. Using R' to route a message around the fault region until the message reach to LN, and then use R.
c. 3) If L_3 have another faults then recursively apply *i*FTDM.
Case 3 (overlap in L_3)
c. 1) 1st copy contains destinations on L_1 using R
c. 2) 2nd copy contains destinations on L_3. Using R' to route a message around the fault region until the message reach to LN, and then use *column path* routing.
c. 3) If L_3 have another faults then recursively apply *i*FTDM.
[d] Repeat the above steps until each destination in the message header is reached.

5. Results and performance analysis

*i*FTDM uses the same simulation study used in FTDM, in which F is used to represent the number of fault regions, R is the number of rows, and C is the number of columns. This configuration creates different networks with a number of processors ranging from 100 to 1080. The size of fault region ranges from 2×2 to 20×20, using 25 destinations. The number of fault regions rangs from 1 to 10, using 30 destinations.

5.1. Latency steps and Traffic steps analysis

Two important performance metrics in direct networks, Network latency steps (NLS) and network traffic steps (NTS), are calculated in this subsection. NLS is the maximum number of channels which the message takes to reach its destinations. NTS is the total number of channels used to send the message to all destinations. They affect the overall performance of the multicomputer system and the granularity of parallelism that can be exploited from the system [13].

Now, NLS and NTS are calculated for *i*FTDM and FT-cube2 routing algorithms. The following formulas can be used to calculate NLS for *i*FTDM. By using the same notations and calculations on [1], we have:

NLS of *i*FTDM, *i*FTDM _Latency, is given by:

$$i\text{FTDM\_Latency} = \text{Max}(LP_{(i)}, RP_{(i)}, L_4) \quad (1)$$

NTS of *i*FTDM, *i*FTDM _Traffic, is given by:

$$i\text{FTDM\_Traffic} = \text{Traffic}_{(i)} + L_4 \quad (2)$$

NLS of FT-cube2, FT_Latency, is given by:

$$\text{FT\_Latency} = \text{Max}\{ \text{Flat}_i, \quad R\ i\ R\ |D|\} \quad (3)$$

Where $\text{Flat}_i = x_{di} - S_x + y_{di} - S_y + 2 * y_{ei} - y_{bi}$

NTS of FT-cube2, FT_Traffic, is given by:

$$\text{FT\_Traffic} = \sum_{i=1}^{|D|} \text{Flat}_i \quad (4)$$

5.2. Network latency steps and Network traffic steps results

By using equations from 1 to 4, NLS and NTS for both algorithms in 2D mesh are

calculated. Figures 4, 5, 6 and 7 show the results. The continuous line represents the results of *i*FTDM, while the dotted line represents the results of FT-cube2.

Figure 4 plots NLS for various values of number of fault regions, ranging from 1 to 10, and |D| is equal to 30. The figure shows that NLS computed by *i*FTDM algorithm decreases as number of fault regions increases, while NLS computed by FT-cube2 algorithm is nearly constant and less than *i*FTDM algorithm. With a small number of fault regions (less than 9), NLS computed by FT-cube2 algorithm is less than that computed by *i*FTDM algorithm, while with a large number of fault regions (greater than or equal 9), NLS computed by *i*FTDM algorithm is nearly the same that computed by FT-cube2 algorithm.

Figure 6 plots NLS for various sizes of one fault region, ranging from 4 to 400 where R x C= 2x2 to 20x20 and |D| is equal to 25. The figure shows that NLS computed by *i*FTDM algorithm nearly constant with sizes between 2x2 and 4x4, then decreases as size of the fault region increases (between 4x4 and 6x6) after that becomes nearly constant. While NLS computed by FT-cube2 algorithm is nearly constant and is always less than *i*FTDM algorithm.

Figure 5 plots NTS for various values of number of fault regions ranging from 1 to 10, and |D| equals 30. The figure shows that, NTS computed by *i*FTDM algorithm is nearly constant (slight increase) as the number of fault regions increases, while NTS computed by FT-cube2 algorithm nearly constant, but greater than with that is calculated by *i*FTDM algorithm.

Figure 7 plots NTS for various sizes of one fault region, ranging from 4 to 400 where R x C=2x2 to 20x20 and |D| is equal to 25. The figure shows that NTS computed by *i*FTDM algorithm is nearly constant (slight decrease) as size of the fault region increases, while NTS computed by FT-cube2 algorithm is nearly constant, but greater than what is calculated by *i*FTDM algorithm. Also, this is because the path computed by FT-cube2 algorithm circles around the fault region.

In all tested cases, NTS computed by *i*FTDM algorithm is less than that computed by FT-cube2 algorithm.

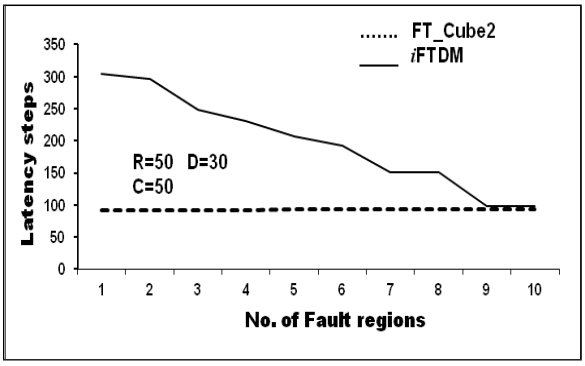


Figure 4. Latency Steps Vs. No. of Fault regions

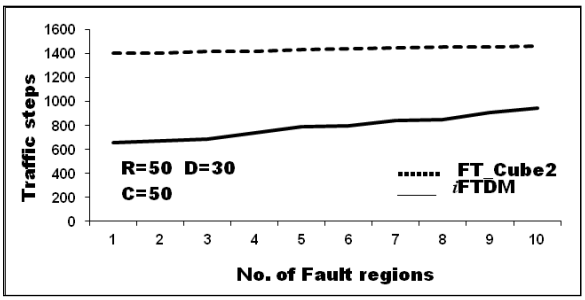


Figure 5. Traffic Steps Vs. No. of Fault regions

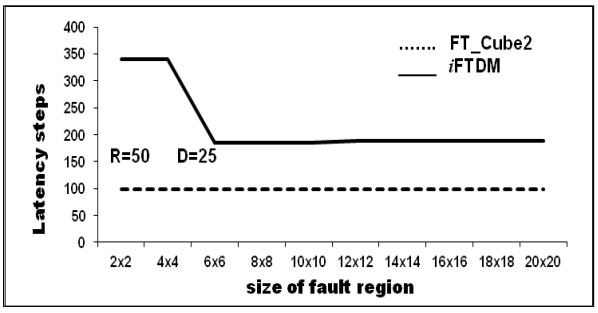


Figure 6. Latency Steps Vs. Size of fault region

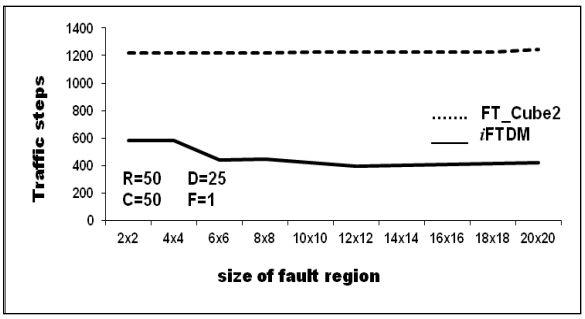


Figure 7. Traffic Steps Vs. Size of fault region

6. Conclusions

In this paper, an efficient fault-tolerant multicast routing algorithm for 2D mesh, *i*FTDM has been proposed, which is an improvement of FTDM algorithm. The proposed algorithm, *i*FTDM, can tolerate convex faults with presence of large number of fault regions and large fault region size. The *i*FTDM tolerates f-chains in meshes with overlapping of convex fault regions. This algorithm is deadlock-free. Simulation results show that *i*FTDM routing algorithm tolerates multiple faulty blocks, convex and concave fault regions, using no virtual channels, and has better performance than FT-cube2 in terms of network traffic steps.

Future work includes applying *i*FTDM on concave fault region with overlapping and adding another two performance metrics in mesh networks, namely network latency time and network traffic time.

7. References

[1] Masoud E. Shaheen, Ahmed Abukmail. "A Fault Tolerant Deadlock-Free Multicast Algorithm for 2D Mesh Multicomputers. " The Journal of management and Engineering Integration (JMEI), Vol. 5, No. 2, 2012.

[2] J. Duato, "A new theory of deadlock-free adaptive routing in wormhole networks," IEEE Trans. Parallel Distrib. Syst., Vol. 4, No. 12, 1993.

[3] Yusuke Fukushima, Masaru Fukushi, Ikuko Eguchi Yairi, Takeshi Hattori, "A Hardware-Oriented Fault-Tolerant Routing Algorithm for Irregular 2D-Mesh Network-on-Chip without Virtual Channels, " Proc. of International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'10), 6-8 Oct., pp. 52-59, Kyoto, Japan, 2010.

[4] Bin Dai, Huabiao Lu, Zhigang Sun, Ziming Song, Yanpeng Ma, Jinshu Su, "MORT: A Technique to Improve Routing Efficiency in Fault-Tolerant Multipath Routing, " MSN, pp.430-435, 2009.

[5] Arshin Rezazadeh, Mahmood Fathy, Amin Hassanzadeh, "A Performance-Enhancing Fault-Tolerant Routing Algorithm for Network-on-Chip in Uniform Traffic, " Asia International Conference on Modeling and Simulation, pp. 614-619, 2009.

[6] A. Rezazadeh, M. Fathy, Gh. Rahnavard, "An Enhanced Fault-Tolerant Routing Algorithm for Mesh Network-on-Chip, " Int. Conf. on Embedded Software and Systems (ICESS 2009), pp. 505-510, 2009.

[7] A. Mejia, J. Flich, J. Duato, S. Reinemo, and T. Skeie. "Segment-Based Routing: An Efficient Fault-Tolerant Routing Algorithm for Meshes and Tori, " IEEE International Parallel & Distributed Processing Symposium (IPDPS), April 2006.

[8] H.-H. Chang and G.-M. Chiu, "An Improved Fault-Tolerant Routing Algorithm in Meshes with Convex Faults", parallel computing vol. 28, pp. 133-149, 2002.

[9] R Libeskind-Hadas, Dominic Mazzoni, and Ranjith Rajagopalan, "Tree-Based Multicasting in Wormhole-Routed Irregular Topologies," in Proc. of the Merged 12th Intl. Parallel Processing Symp. And the 9th Symp. On Parallel and Distributed Processing, Orlando, Florida, Apr. 1998.

[10] K. Hwang, Advanced Computer Architecture: Parallelism, Scalability, Programmability, McGraw-Hill, New York, 1993.

[11] R.V. Boppana, S. Chalasani, and C.S. Raghvendra, "Resource Deadlocks and Performance of Multicast Routing Algorithms,"IEEE Trans. on Parallel and Distributed Systems, vol. 9, no. 6, June. 1998.

[12] X. Lin and L.M. Ni, "Deadlock-Free Multicast Wormhole routing in Multicomputer Networks," In

Proc. of the Intl. Symp. On Computer Architecture, pp. 116-124, 1991.

[13] M. E. Shaheen, "High Distributed-Memory Systems Performance," master thesis, Computer Scince Dept., Cairo University (Fayoum Branch), 2005.

Application of Six Sigma Methodology to Reduce Environmental and Economic Impacts of Discharging the Produced Water in Southern Iraqi Oil Fields

Maher Al-Shamkhani and Ahmed K. Elshennawy
University of Central Florida
maher_t@knkights.ucf.edu, Ahmad.elshennawy@ucf.edu

Abstract

Produced Water (PW) represents the largest volume of waste that is normally generated during oil and gas production. It has large amounts of contaminants that can cause undesired environmental and economic impacts. Current methods for the management of produced water rely heavily on the types and concentrations of contaminants, which are field dependent and can vary from one oil field to another. Produced water can be converted to fresh water if these contaminants are removed or reduced to the acceptable drinking water quality level. Removing dissolved hydrocarbons and toxic compounds from produced water may eventually lead to increasing oil production rate and reducing amounts of discharged harmful contaminants. In order to identify the types of these contaminants, effective tools and methods should be used. The purpose of this paper is to introduce and implement six sigma methodology in oil and gas industries. The paper illustrates that, as a result of implanting six sigma, negative impacts of produced water discharges have been reduced.

1. Introduction

With increasing demand and consumption of oil, the frequency of petroleum-related ecological incidents is increasing. In 1949, the Basrah Petroleum Company in Iraq discovered the Zubair field, which is located in Southern Iraq. It has been recently considered one of the largest oil fields in the world. Currently, it is holding around 4.1 billion barrels of crude oil [1]. In 2009, the Eni Company won the service contract for that field, and an expansion program is taking place in order to develop the infrastructure of the Zubair field. As a result of this program, the production of oil is expected to increase from 195,000 to 1,125,000 BPD (Barrel Per Day) by 2017 [2]. In addition, more than 200 wells will be drilled. Furthermore, treatment facilities, required collection network for the fields, and the reconstruction of the existing plant will be accomplished by the end of this program. Since the volume of Produced Water has increased from 4,000 BPD in 2008 to 35,000 BPD in 2012, production volume is expected to increase to more than 1,169,000 BPD in the near future [2]. This excessive amount of produced can be a source of fresh water if properly treated and can prove beneficial to humans who are living closer to the Zubair field or to the oil field itself. Furthermore, if this large amount of PW is properly managed and effectively treated, the reinjection process into oil wells can be achieved [3], [4],[5], [6]. Produced water has various contaminants such as heavy metals, sands, dissolved gases, bacteria, and dissolved hydrocarbons and the current method that is managing this contaminated water is not

properly managed and thus has negative environmental and economic impacts on the Dammam formation, Zubair aquifer, employees, and human resources in the areas surrounding that field. If properly treated, the productivity of oil and fresh water resources will increase, and the negative environmental and economic impacts will decrease. In order to achieve these results, an effective management method of the produced water in that field is needed. The issue now is to develop an effective way of reinjecting produced water into Zubair oil wells and reusing it as a fresh water resource.

In this paper, we introduce and implement six sigma methodologies to identify the root causes of having high concentrations of contaminants in produced water and to recommend new technology and management method for its treatment prior to its reuse or disposal.

2. Degasing Station Process Description

The South Oil Company (SOC) has been using Degassing Stations (DS) to separate oil from gas. These stations consist of dehydrator and desalting units (desalters), which are used to accomplish the separation process of oil, gas, and formation water. The dehydrator unit is responsible for separating the oil from formation water by which almost 98% of water can be separated. The desalter is responsible for washing the dehydrated mixture from the salt in which fresh water is required to accomplish the desalting process. Currently, this wash water is obtained from a river that is called Garmat Ali River which is about 10 miles away from the degassing stations. An excessive amount of produced water has been produced with oil and gas production activities in that field. This water has been managed by pumping it into naturally evaporating ponds (NEPs), see Figure 1. By using these ponds, most of water will be lost and evaporated to the environment and large accumulated solid and liquid wastes can be observed in these ponds and near to the discharge points [7].

3. Six Sigma Approach

In this study, the Six Sigma approach is implemented to evaluate PW stream which is currently discharged from the DS of Zubair oil field. This approach provides effective tools that can be used to: 1) identify the most hazardous compounds in the discharged produced water, 2) measure the amount of these hazardous compounds, and 3) identify the root causes of having high percentage of such compounds in the discharges. Analyzing these causes and relating them to the current management method of produced water in the selected locations can help to successfully select new management and technology that can sustain for the long term with current specification of produced water and current operation and production conditions. Therefore, developing new policy changes for the current management method of produced water and proposing a new treatment technology that can be used to reduce the amount of contaminants into acceptable levels prior to its discharge or reuse are added values of this study.

We will follow the traditional DMAIC methodology of six sigma, that stands for the five phases of process improvement known as: Define, Measure, Analyze, Improve, and Control phases.

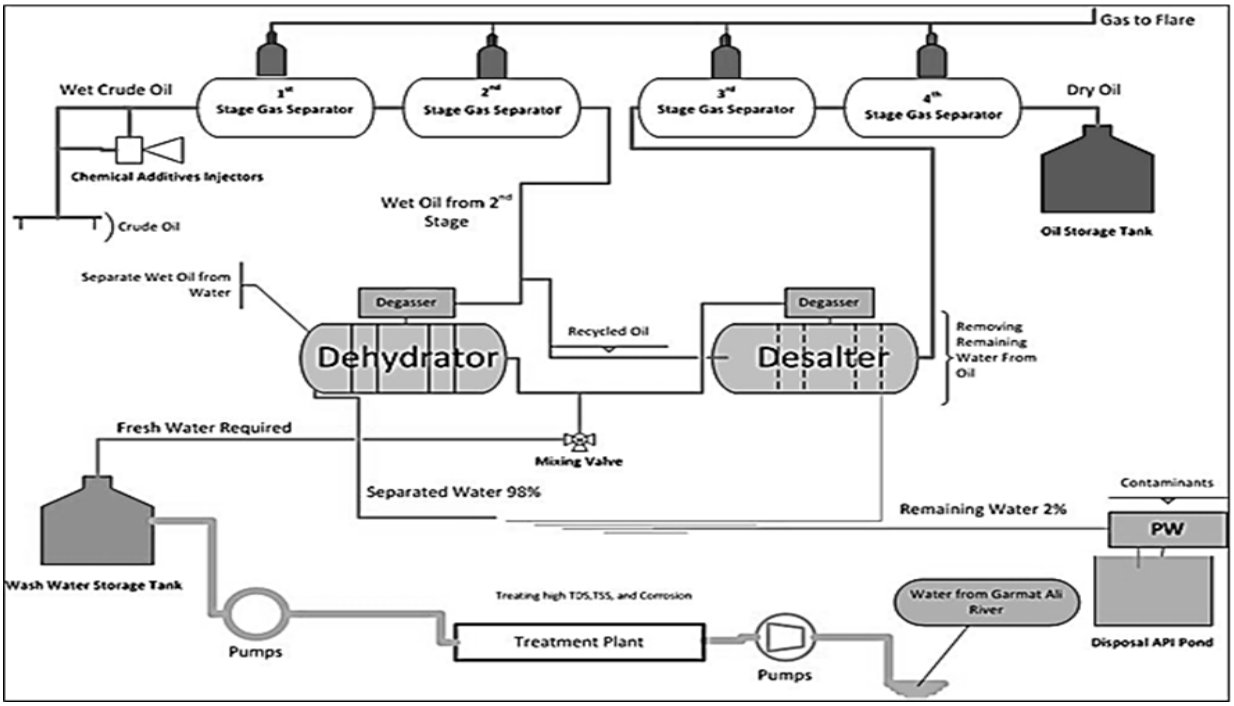


Figure 1. A Schematic Diagram for DS

3.1 Define Phase

In this phase of the DMAIC approach, the limitations of the project and its current and future benefits are identified. The customer requirements, which are known as Voice of Customer (VOC), are determined in order to set the best tools and methods that could be used to meet or exceed the customers' expectations [8]. Furthermore, important customers' needs, known as Critical to Quality (CTQs), are also defined and are used to set the best targets and to use proper tools in order to meet or exceed these expectations [8], [9].

3.1.1. Stakeholders Analysis

Stakeholder analysis is performed to distinguish between vital, supportive, and adversarial stakeholders. This analysis is performed in the early stages of the DMAIC approach because it is important to know who will be a supportive for the initiatives toward problem-solving and quality improvement stages and who is going to be a potential adversarial to these initiatives. For the purpose of the stakeholders' identification, the Stakeholders Analysis Matrix (SAM) is used, see Table 1. The results obtained from the SAM are demonstrated by using interest/ power plot and attitude/ activity plot; see Figure 2 and Figure 3 below.

In Figure 2, the reference line represents the ideal balance for a vital stakeholder. Points above the line represent stakeholders with potentially high influence on the success of the project; they can be either powerful supporters or powerful detractors. In Figure 3, the reference line on the left marks the point at which stakeholders are considered potentially adversarial to the project.

Table 1. Stakeholders Analysis Matrix

Stakeholder Categories	Relevant Stakeholders	Code	Attitude (-10-10)	Activity (0-10)	Attitude Rating	Power (0-10)	Interest (0-10)	Power Rating
Environmental Protection Agency	EPA	EPA	8	9	72.00	5	9	45.00
South Oil Company	SOC	SOC	5	10	50.00	10	6	60.00
Coworkers	Engineers and Workers	HU	9	7	63.00	5	2	10.00
Current unit Plants	Operators	OP	-6	9	-54.00	8	10	80.00
Current Production Rate	Producers	PR	-10	9	-90.00	8	1	8.00
Wells Management	Managers	MA	-9	6	-54.00	8	7	56.00
Governors	Ministry of Oil	GOV	9	10	90.00	7	10	70.00
Contractors	International Oil companies	F.CO	9	8	72.00	8	10	80.00
Domestic Governors	Do.Gov	D.GOV	-6	8	-48.00	7	7	49.00

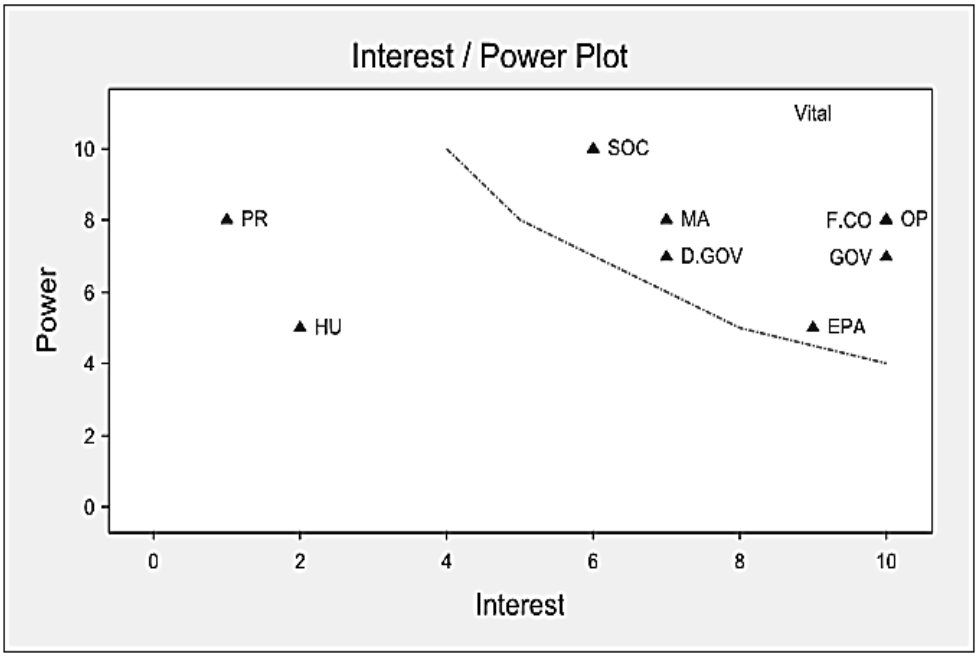


Figure 2. Interest/ Power Plot

Points to the left of this line represent stakeholders that could present roadblocks. The reference line on the right marks the point at which stakeholders are considered potentially supportive for the project. Points to the right of this line represent stakeholders that could provide assistance in overcoming roadblocks.

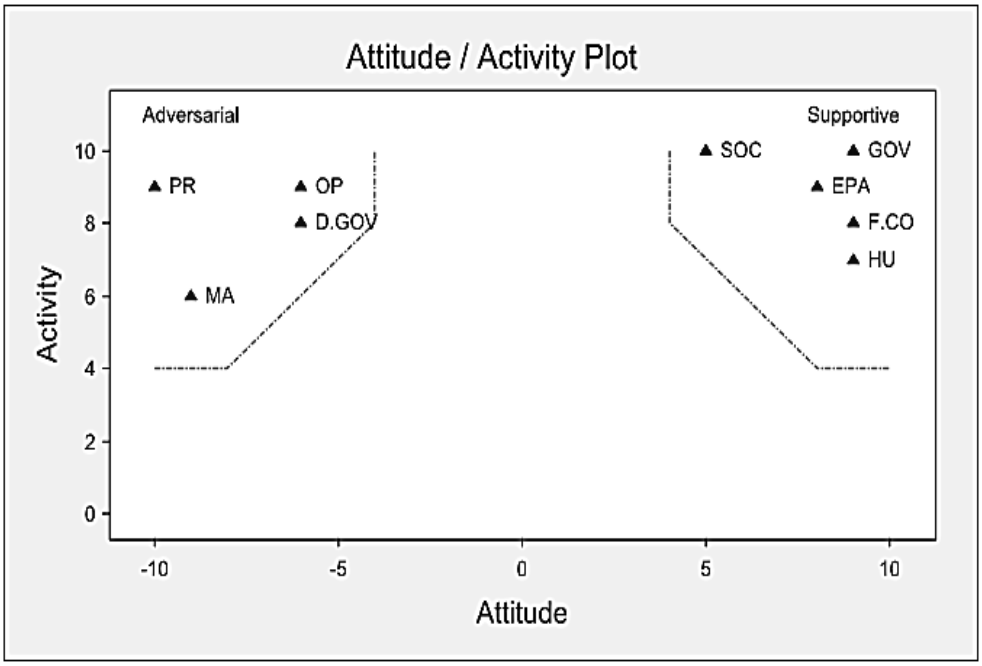


Figure 3. Attitude/ Activity Plot

3.1.2 SIPOC

The SIPOC process which stands for Suppliers, Inputs, Process, Outputs, and Customers is used to explain what and who was involved in the study. Defining the customers and the sources of information that are used in the next phases is also demonstrated by using SIPOC. The start and the end point for each involved process were also defined in this section.

.

Table 2. SIPOC

S	I	P	O	C
Suppliers	**Inputs**	**Process**	**Outputs**	**Customers**
Provider	Input requirements and measurments	**Description**	Output requirements and measurments	Receiver
Petrochemical Labs	Sampling and Testing PW & Oil	Performing Chemical and Physical Tests	Valid Oil and PW Specifications	Measurment & Control Department
	PW Physical & Chemical Characteristics	Determining and Identifying Contaminants in PW	Types and Amount of Contaminants in PW	Zubair field Management
Zubair field Management	Oil and PW Characteristics Reports	Making a Proper Decision to Reduce Discharge Rate of PW	Reduce Environmental and Economic Hazards	SOC

SUPPLIERS

Chemists in petrochemical labs were performing physical and chemical tests for both oil and produced water samples for this study, taken from the output stream of the dehydrator units at different locations. Also, they were reporting types and concentrations of contaminants in produced water.

Zubair field management departments were responsible for reviewing the routine operations and production reports that could help measure the effectiveness of oil-gas production operations and separation equipment.

Furthermore, these departments were responsible for supporting and providing all needs of production, maintenance, sales, and other departments, such as, purchasing required equipment and parts for operation, production, maintenance, and control processes.

INPUTS

Samples from oil, sludge, formation, and produced water were taken and tested by petrochemical labs in order to study physical and chemical properties.

Results obtained from the petrochemical labs were included in this study.

Reports that discussed physical and chemical properties of discharged produced water from the Zubair oil field were also included in this study.

PROCESS

Performing chemical and physical tests for produced water samples.

Evaluating the main contaminants in the discharged PW and identifying its environmental and economic impacts.

Improving the current method for managing PW in the Zubair field.

OUTPUTS

The ultimate output was reducing the environmental and economic impacts that can result from discharging contaminated produced water into areas surrounding Zubair field. This output could be achieved if produced water properties were determined accurately, and that would help to select the best method for managing that water.

CUSTOMERS

The internal customers in this project were measurement and control department of South Oil Company, workforces, and the Zubair oil field management departments. From the perspective of safety, conducting chemical and physical tests could help measure the harmful contaminants which must be eliminated or controlled to protect employees during handling of that water. Managing produced water properly could help protect the Dammam formation and Zubair aquifer, and that could also help protect the environment in southern Iraq. Eliminating or at least reducing the amount of contaminants associated with produced water to the acceptable levels could improve the protection of humans who are living close to the Zubair oil Field areas who are considered external customers.

3.1.3 Process Flow Chart

In order to understand the current basic processes that were involved in producing oil and contaminated produced water in Zubair field, the flow chart in Figure 4 is developed.

3.1.4 The Voice of Customer (VOC)

Understanding the needs for both internal and external customers is required. In this paper, several approaches were used in order to their requirements and expectations.

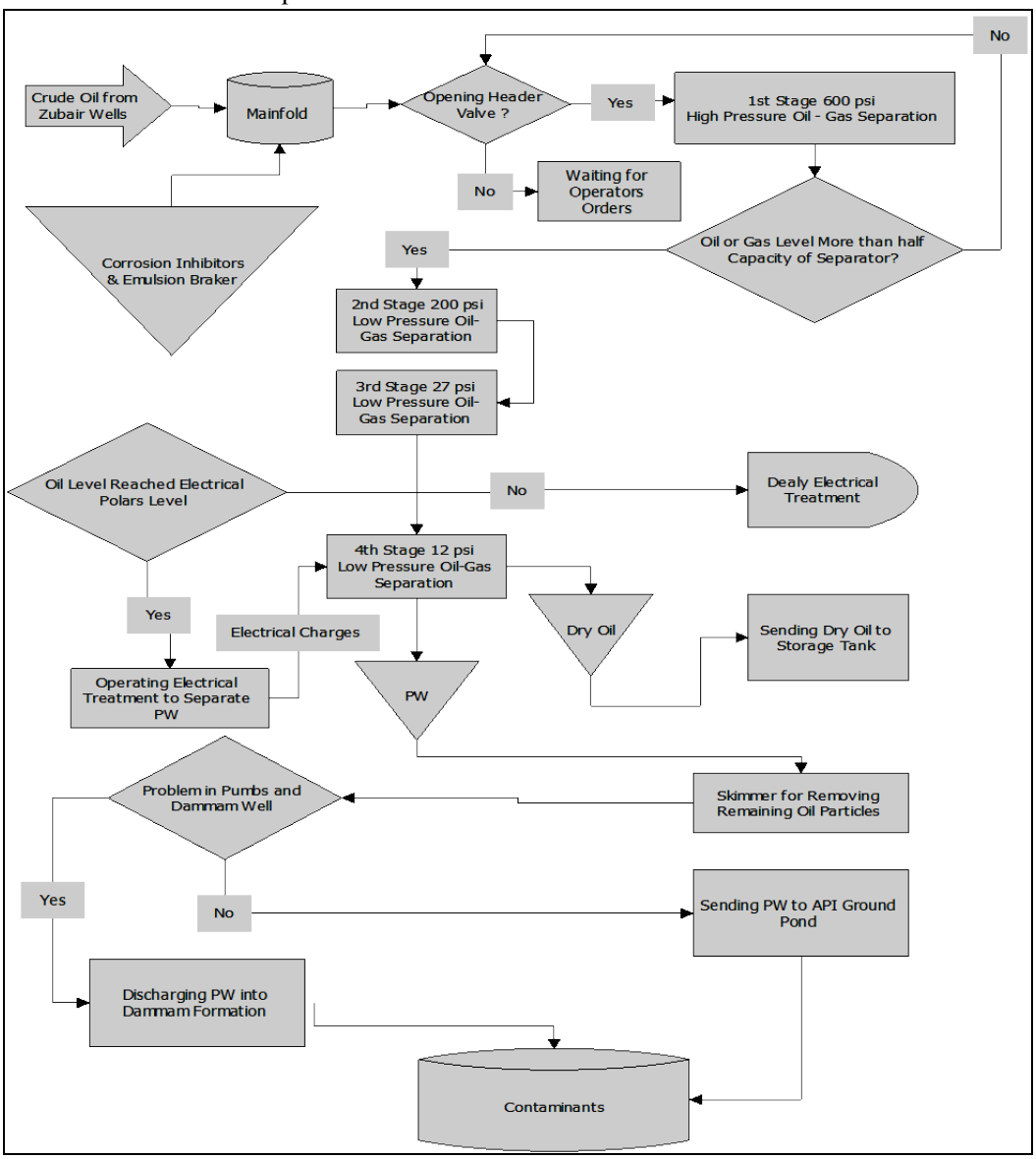


Figure 4. Flowchart of DS Processes

Direct customer contact, internet exploring, and field intelligence are used to gather information about SOC and its requirements that are related to main concerns about increase in the discharge rate of produced water with high concentrations of harmful contaminants. South Oil Company requires having an effective produced water treatment plant to treat PW prior to its reuse.

3.1.5 Key process input variables and key process output variables

For this study, it is important to determine the key process output variables, known as

KPOVs. These are the factors that could influence the amounts of contaminants in produced water (PW) that are required to be identified. For this study, our KPOV is considered to be the increase in the amount of contaminants in produced water (PW).

In addition, we identified the key process input variables (KPIVs) as:

- PW Management Method
- PW Analysis Methods
- Operation and Production Methods
- Operation and production Plants
- Field Observation and Control Methods
- Maintenance Methods
- Information Technology Method
- The Nature Causes

3.2 Measure

In this phase of the DMAIC approach, the internal processes and activities that could impact the Critical to Quality characteristics (CTQs) are measured. The KPIVs and KPOV are discussed in details in order to measure the relationships amongst them and with identified CTQs. Identifying types and quantities of different contaminants in PW were helpful to determine the root causes of corrosion, equipment failure rate, and the high amount of normally occurring radioactive materials (NORM) and toxic materials in the discharged PW.

3.2.1 Critical to Quality (CTQ)

In order to capture the VOC and CTQs in a more detailed mode, the Quality Function Deployment (QFD) approach is used to determine the relationship between customers' needs and technical requirements that were needed to meet or exceed customer expectations. The principle focus of QFD is identifying the customer needs properly in the early stages, and that would help to reduce waste, such as rework, redesign, and rethink to find the quick solution that will be another problem in the near future [8]. The Technical Importanc Rating (TIR) was determined to identify which requirments have high weights and need to set high targets that could help to meet critical needs of customers. Internet monitoring for all local and foreign tenders that have been requested by Iraqi ministry of oil and the SOC was helped to gather more details about CTQs; and that also helped to identify the most important technical factors that could impact the CTQs. The technical importance rating was very high and critical for each of the following requirements:

1. Best Approach for Managing PW
2. Construct PW Treatment Plant
4. Construct PW Reinjection Units
6. Construct Geochemistry Labs

The results obtained from both VOCM and QFD indicated that the main requirements are related to manage, treat, reuse, and/or reinject PW. As a result, the CTQs are identified and represented the important needs of customers as follows:

Finding the best method for Managing PW

Converting PW to usable water

It is important to notice that there is a strong relationship between these two requirements. Converting PW to usable or clean water is the most important requirement because the management method of PW is highly based on the quality of that water.

3.2.2 Key Process Output Variables Measurement

In the define phase, increasing in the discharge rate of the PW is illustrated, but types and concentrations of contaminants in PW are not measured. Therefore, in this section, types and characteristics of these contaminants are measured and explained by testing PW samples

that were taken from the output stream of four dehydrator units, which are located at different locations in the Zubair oil field. These samples are tested in 2012. Physical and chemical specifications of PW are summarized in Table 3. In addition, average values of most important properties of PW before and after treatment are listed in Table 4.

Table 3. PW properties at different locations

Sample Point Location	Dehydrator Alzubair	Dehydrator Alzubair Musharif	Dehydrator Hammar	Dehydrator Hammar Musharif
Density at 60 F	1.125	1.145	1.13	1.13
Temperature C	60	60	47	52
Condictivity micro/cm	341000	388000	365000	356000
TDS mg/l	170000	195000	184000	178000
TSS mg/l	180	136	82	165
Turbidity NTU	126	116	91	96
PH	6.13	5.92	6.1	6.03
Chlorine	0	0	0	0
Chloride	115730	133480	127800	122120
Sulphate	48201	52737	48491	30289
Phosphate	<1	<1	<1	<1
Bicarbonate	244	244	244	244
Calcium	3920	7120	6240	5920
Magnesium	1749	1263	1069	729
Sodium	74980	86480	82800	79120
Iron Content	70	50	90	50
Zink	<2	<2	<2	<2
Salinity	190710	219960	210600	201240
Nitrate	<1	<1	<1	<1
Salt %	180	200	190	180
Hardness Ca	9800	17800	15600	14800
Hardness Mg	6069	4382	3709	2529
Total Hardness	15869	22182	19309	17392

Table 4. PW Properties before and after Treatment

Before Treatment		After Treatment	
PH	5	PH	6.5-7.5
TSS	300 Mg/l	TSS	<2 Mg/l
Turbidity	700 NTU	Turbidity	unknown
Particle Size	60 Micro-m	Particle Size	<4 Micro-m
TDS	250,000 Mg/l	TDS	250,000 Mg/l
Oil and Grease	1,000 Mg/l	Oil and Grease	<5 Mg/l
Total Iron	300 Mg/l	Total Iron	<5 Mg/l
Dissolved O2	>2 Mg/l	Dissolved Gases	<0.02 Mg/l
Bacteria	All Types	Bacteria	None
CO2	470 Mg/l	CO2	<0.02 Mg/l

3.2.2.1 KPOV- Radon Concentration

Sludge and formation samples were taken from different locations at the southern oil fields and tested for radon concentration. Pareto chart was used to show the average of radon concentrations at different selected locations [10] as illustrated in Figure 5.

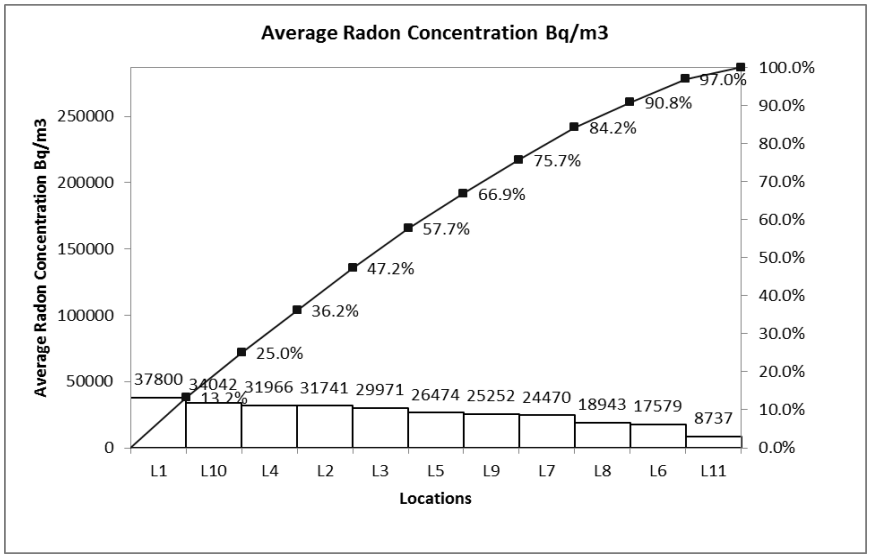


Figure 5. Pareto Chart for Average Radon Concentration

L1 is the location of sludge samples that have been tested for radon gas concentration and that

have been taken from the DS of the southern Rumaila oil fields. The probability of getting cancer is $1.6x10^{-3}$ for each 37 Bq/m^3 [11], thus from the obtained average which is 26089 Bq/m^3, the probability increases to 705 multiple[10].

3.2.2.2 KPOV- Sodium Adsorption Ratio

In their work, (Dallbauman and Sirivedhin, 2005) formulated an equation that can be used to find the Sodium Adsorption Ratio (SAR). SAR value equals the ratio of the sodium concentration to the square root of the average for both calcium and magnesium concentrations respectively as in the following:

$$SAR \quad \frac{[Na^{2}]}{\sqrt{\frac{([Ca^{2}] \quad [Mg^{2}])}{2}}}$$

By using this equation, it can be determined whether the PW will have future negative impact if it used in an irrigation process. In 2006, the American Petroleum Institute (API) found that decreasing in soil permeability and increasing susceptibility to erosion associated with irrigation water which has SAR value > 6. The results obtained from the calculations of SAR at the selected dehydrator units, it was reasonable to conclude that the SAR values were very high and it exceeded the normal expected value as shown in Table 5.

Table 5. SAR Calculated Values

Dehydrator	SAR Calculated
Zubair	249.39
Zubair Musharif	247.6
Hammar	254.21
Hammar Musharif	257.56

3.2.3 Key Process Input Variables Measurements

The amounts of scales, precipitated and corrosive materials, and NORM are observed in pipes, valves, pumps, storage tanks, and other field facilities. Therefore, ineffective monitoring, cleaning, and maintenance plan helped to increase these amounts, then the PW, oily sludge, and other disposals have high concentrations of these risky contaminants and that need to rethink again about type of management method used to handle, transport, and dispose them safely. The root causes of high amounts of various contaminants in PW are twofold: a) current management method of discharging PW and b) other causes that are attributed to existing of various contaminants in the oil fields equipment, pipes, and all activities that could help to increase the amount of these contaminants, such as ineffective maintenance plan and type of chemicals used during oil and gas production operations and personal activities.

3.3 Analyze Phase

In this phase of the Six Sigma project, conducting a problem solving approach is performed to determine why there high concentrations of contaminants in PW, and how the current management method of PW in Zubair oil field was contributing to the high values of these contaminants. At the Analyze phase, we focus on analyzing the following:

The main sources of contaminants in the DS of the Zubair oil field.

The current method for managing PW in the Zubair oil field.

Corrosion

Having high iron content (IC) in any pipe system, equipment, or plant causes a high rate of corrosion that could cause an increase in the equipment failure rate and amounts of corrosive

materials, scales and deposits. From the main specifications of PW before treatment, the average of high IC was more than 300 mg/l, which was very high. The reason behind that was the multiple corrosions during oil operation and production processes and types of corrosion inhibitors or chemicals used. Causes of increasing the corrosion rate are analyzed in details as described in the Fishbone diagram of Figure 6.

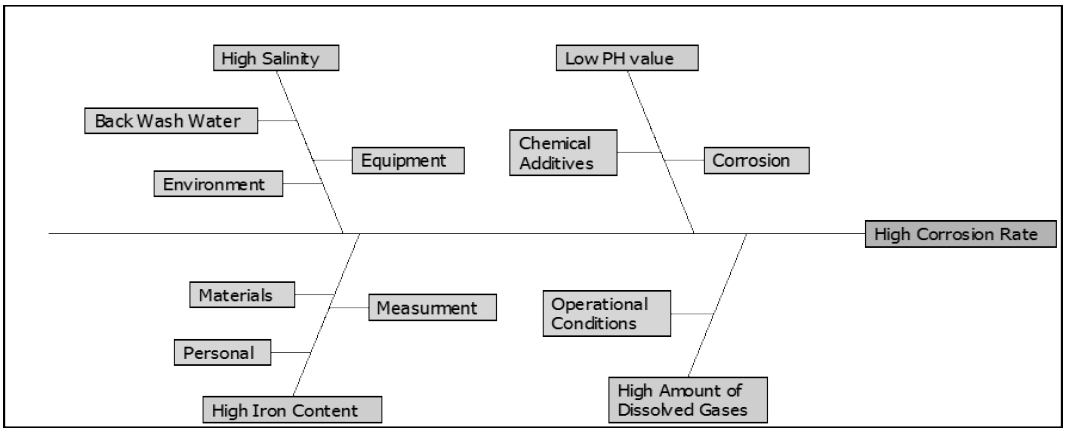


Figure 6. Fishbone Diagram for the Corrosion

After identifying the root causes of having high corrosion rate, we used the system thinking approach (STA) to demonstrate the relationship between these causes. Figure 7 shows the factors that contribute to the increase of the amount of corrosive materials.

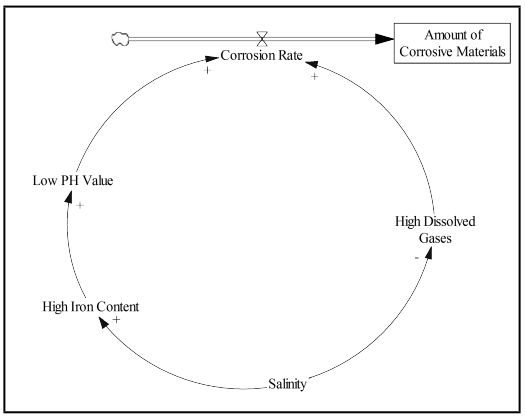


Figure 7. Causal loop Diagram for the Corrosion

Field Equipment

One of the keys that can affect the concentrations of the contaminants in the discharged PW is the efficiency of the equipment that is currently existed in the DS of Zubair oil field. It was determined that the following factors are contributors:

- Old wet crude oil separators are used
- Power supply and power cables failures are observed
- Failures caused by control valves
- Old pipe systems that contain accumulated deposit and precipitated solids (Scales).

Personnel and laboratories

We have identified the following as reasons that contribute to the weakness of the current management system of produced water:

- Lack of communication between internal and external customers
- Lack of data sharing between oil wells and management departments
- Lack of information about Iraqi oil

fields, in particular those related to PW problems.

Ineffective maintenance plan

Lack of training in PW management

Unsafe maintenance and observation areas surrounding southern oilfields due to the existence of normally occurring radioactive materials (NORM)

Failure detection plans and/or failure prevention plans are not present.

Chemists and geologists in the geochemistry and petrochemical labs rely on the predication more than performing real tests. Laboratories lack the existence of precise and efficient oil and PW test equipment.

Other Causes

Producing high concentrations of radioactive materials during oil production

Existing high amounts of heavy metals in underground geological formation

High salinity of formation water

Current Management Method

Since the amount of PW has increased enormously with increase of oil production, SOC has been injecting PW into naturally evaporating ponds (NEPs) and disposing it into Dammam formation in order to keep normal production of oil. Discharging large amount of PW into NEPs was considered the classic treatment method or quick fix for PW problem and that was represented by evaporating PW by solar energy. This kind of treatment is not needed to use chemicals and energy, but waste disposal is required for the accumulated materials. Because the production of oil is expected to be increased, the number of NEPs is also required to be increased. In fact, discharging PW without treatment requires large and safe discharging areas to ensure that contaminated water will not pollute or affect the environment and humans. If PW discharged into surrounding areas without treatment, formation plugging for that area could occur, so the permeability of that formation will decrease. Large disposal area and formation plugging can be considered main constraints that can cause delay in oilfields development projects.

In addition, discharging PW with high concentrations of chemicals can cause corrosion and acute toxicity increases near to the discharge points. Furthermore, high corrosion rate can increase equipment failure rate. Therefore, more efforts are required to fix problems that have been associated with oil production and resulted from discharging contaminated PW. These efforts can be represented by hiring experts or large oil companies, which have good experiences in managing PW. That was exactly what SOC did in order to improve oil production and current state of the Zubair oil field. Because PW has two main constituents, which are oil and grease particles, the amount of these particles can be added to the current amount of oil that has been produced if it separated from PW. As a result, the opportunities of increasing oil production in Zubair filed will increase. In fact, this amount of fresh water is provided from Garmat Ali River after reducing amounts of TDS and TSS. Furthermore, anti-corrosion additives are normally used to reduce the corrosion impacts on drilling equipment, operation units, and production plants. Because of the high salinity in BWW (Back Wash Water) which is exceeded 200,000 ppm, the treatment is expected to be very expensive. Supplying BWW from Garmat Ali River to the Zubair field requires long distance pumping units, long pipe systems that will reach to more than 10 miles away from DS. These systems and plants require operational, control, and maintenance cost.

Converting PW to fresh water will help reduce the required amount of supplied water from Garmat Ali River. Furthermore, the fresh water can be injected again into oil wells to maintain oil wells pressures, and then increase oil productivity of these wells.

3.4 Improve Phase

In this phase, reducing the amount of contaminants that have being discharged with PW is the main objective. The results obtained from previous phases are used to develop an effective framework that could be used to manage PW in the Zubair oilfield effectively. These results are also showed that there is a strong relationship between the quality for that water and type of management and treatment technologies. Since the current method is considered ineffective, continuous improvement initiatives toward finding the best method and effective technology to treat that water, to decrease the amount of contaminants, and to convert PW to usable water are started by proposing stationary PW treatment plant.

3.4.1 Produced Water Treatment Plant

First of all, from the analyze phase, one of the most causes of high salinity and corrosion problems were related to use of BWW that has been treated and supplied from Garmat Ali River through old pipe systems that contain scales and deposits. A treatment plant for that water closer from that river was existed and it has been used to reduce TSS from more than 200 ppm to 3 ppm with particle size equals 10- micron. Furthermore, anti-corrosion additives have also been used as a part of this treatment with pesticides to kill different kinds of bacteria. However, once that water pumped through the transportation pipe system, an intermediate contact with deposits, corrosive materials, and scales could occur. As a result, these cumulative contaminants could be carried with BWW into the DS. Therefore, the amount of these contaminants can be added to those that were coming with formation water during oil extraction operations. Thus, the concentration of contaminants that could be associated with the PW, which was being discharged from the dehydrator and desalter units, was increased.

Secondly, existing of PW treatment plant at DS has some benefits. Firstly, treating PW closer from CDS will help to reduce the demand on obtaining water from Garmat Ali- River. As a result, the amount of moving suspended scales and deposits in pipe systems from that river to the CDS will decrease. Secondly, converting PW to usable water will help to use it not only for production operation purposes, but also, it could be used for cleaning field facilities, such as, cleaning drilling equipment or can be used as a cooler fluid for some cooling systems.

Furthermore, if PW effectively treated, the reinjection process into oil wells will be possible. The latest method will increase oil production and maintain oil wells pressure. Finally, the disposal rate of PW will be decreased and that will result in decreasing the environmental and economic impacts. Due to the fact that there are different existing technologies for the purpose of PW treatment, it was necessary to search for the best methodology that can help to select an effective treatment technology for that water with less efforts and time. Meeting customer's requirements requires studying different treatment technologies taking into account some important criteria.

3.4.2 Produced Water Treatment Plant Selection Approach

According to VOC and CTQs results that have been measured and analyzed in the previous phases and based on the customer's requirements, decision making tools are used to

select an effective management method for PW. Since current specifications of PW are measured and analyzed, further research is conducted in order to meet the required specifications of PW after treatment (customer's needs). Meeting these requirements could help to convert PW to clean water and that is one of the most important needs of SOC. According to the VOC, PW treatment plant is required to convert PW to usable water. The required specifications of PW after treatment were helped to identify the goals of this study and to select the best PW management technology. Meeting these requirements helped to improve the current state of Iraqi oilfields by reducing negative environmental and economic impacts of PW.

In fact, different technologies for managing that water are available in current markets. However, selecting the best technology with the respect to main important factors, such as cost, environmental, technical requirements, and health and safety are performed by using MCDM (Multi Criteria Decision Making) methodology.

3.4.3 Analytical Hierarchy Process (AHP)

AHP is an emerging solution to complex decision making processes and it is widely used as the best method for making decision in developing an effective strategic plan for organizations and selecting new manufacturing technologies. Since our problem is the increase in the concentrations of contaminants in PW, different alternatives are identified with respect to four basic criteria as follows:

1. Technical Feasibility
2. Cost
3. Environment
4. Health and Safety

Four technologies are selected as alternatives. These technologies were studied carefully, considering the customer requirements and the results obtained from the study. These technologies are already designated and available in technology markets. Each of them has advantages and disadvantages and some of them have been used by some oil industries around the world. These technologies are:

1- Technology A1- Hydro cyclones
2- Technology A2- Media Filtration
3- Technology A3- Membranes Filtration
4- Technology A4- Evaporation pond.

From the above, it is important to notice that the current management method of PW in the Zubair oil field is also selected and included in the alternatives selection process. The reason behind that is to compare it with other selected technologies. Technical reports and published papers are used to investigate how each technology could be used to achieve the desired requirements. The model that is used to select one of the technologies as an optimum solution for PW problem is developed by using AHP methodology and is shown in Figure 8.

3.4.4 AHP model Description

The AHP model is develop by using Superdecision software. This model connects the main goal of this study, which is finding the best method for managing PW effectively in the southern Iraqi oil fields, to the criteria factors as described above. Then, each selected criterion is connected to its sub-criteria. For instance, the criterion Environmental has its own sub- criteria which are ecological risks, solid wastes, liquid wastes, and NORM. Finally, the main goal is connected to the selected alternatives technologies through all these sub-criteria. These connections are allowed to have the whole model connected to perform pairwise compassions between these technologies and other criteria and sub-criteria in the model.

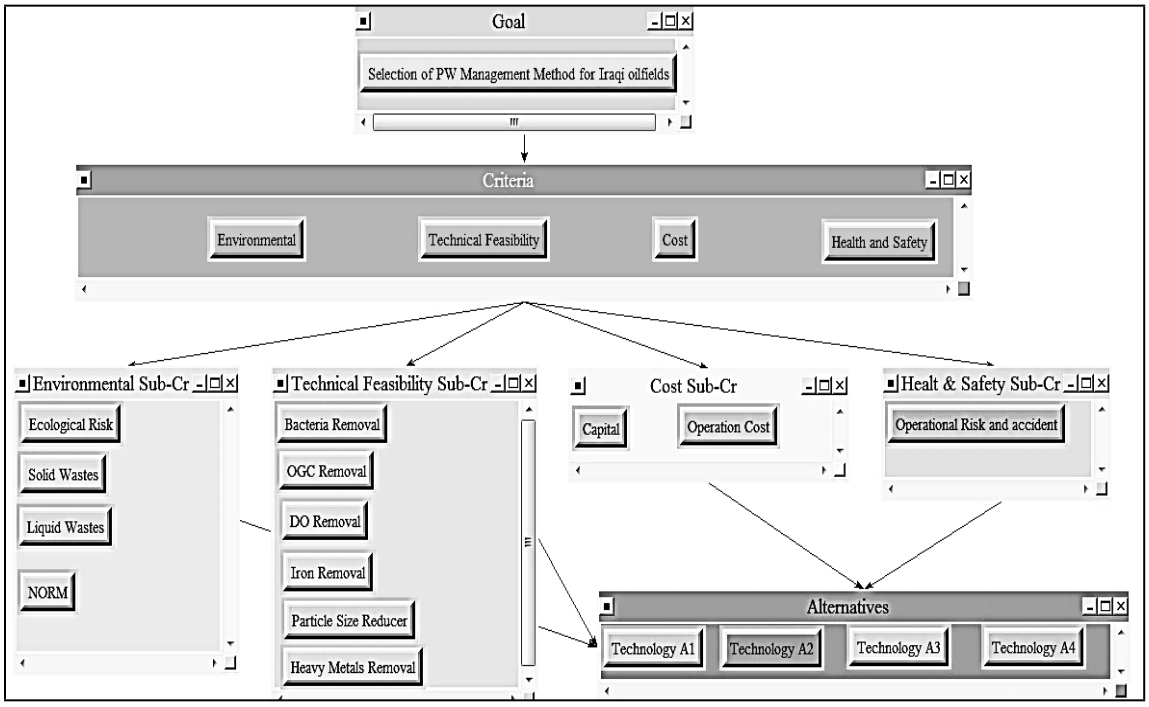


Figure 8. AHP Model

3.4.5 Synthesizing the AHP Model

In this step, the optimum method that could be used to manage PW in Iraqi oilfields was identified. The best way that could be used to report the result was synthesizing the whole model. The results shown in Figure 9 show that the best method is using membrane filtration which is technology A3 with the normalized value equaled to 0.404603. The second alternative technology that can be used for the same purpose is technology A2 (media filtration) with normalized value equaled to 0.243328. Technology A1 is considered an intermediate candidate between the above technologies with normalized value equals 0.208771. Finally, technology A4, which is the current method for managing PW in the Zubair oil field, is considered the bad alternative that cannot be used to achieve the goal of this study with normalized weight 0.143298.

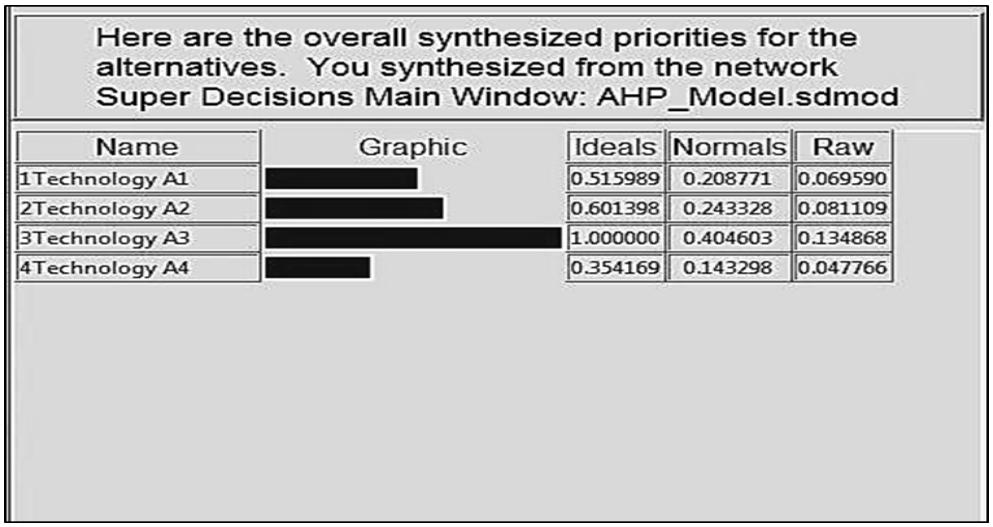


Figure 9. Results of Synthesizing the AHP Model

3.5 Control Phase

This is the last phase of Six Sigma project in which right actions, correct decisions, and failure prevention strategies are provided in the control plan. The control plan could help in identifying the most popular failures or procedures that could affect the sustainability and the performance of the selected technology. If the membrane filtration technology has been selected to treat the PW in the southern Iraqi oilfields, some procedures and processes are required to maintain the high performance of that technology. Therefore, the high level control plan is developed and introduced in this phase to propose some procedures and actions that could be taken to protect the sustainability of the selected technology and to ensure that most of the common causes of problems, such as corrosion problems, types of chemical used could be eliminated or at least reduced for the long term operation period.

4. Conclusion

The application of Six Sigma in oil and gas industries using the DMAIC approach is a powerful method to successfully identify problems, measure and analyze their causes, remove these causes by using quality control tools, improve the current states of existing systems, and control those systems for the long term period. Implementing of quality principles, practices, and tools in the selected case study is effective to identify the main contaminants in produced water and uncover the different causes of identified contaminants. This study introduces new quality concepts, principles, tools, and methods that can be used to solve problems, improve systems, and manage organizations effectively in oil and gas industry in Iraq.

5. Recommendations

Implementing Six Sigma in oil and gas industries can help manage these industries and improve their operations.

Providing training in quality tools and methods to quality control engineering and design departments will help them identify problems, remove their causes, and reduce wastes.

Quality management concepts are very important for top management and engineers who are working in oil and gas industries. These concepts, such as

six sigma, provide them efficient methods and tools whenever initiatives toward quality improvement processes are started within their organizations.

Statistical process control tools and system analysis method are effective to perform data measurement and analysis precisely.

Statistical thinking approach is more beneficial to use whenever root causes of problems and their effects were required to be identified.

For the selected technology, a pilot treatment plant is highly recommended to construct prior to construct the whole treatment system for PW in the southern Iraqi oil fields. Since the dehydrator of Zubair Musharif is discharging high amount of contaminants in PW stream, a pilot treatment plant should be constructed at the degasing station of the Zubair Musharif.

Pilot treatment performance monitoring should be performed by testing samples from filtrated stream to check for TDS, IC, TSS, OIWC, existing of bacteria, and NORM.

Regular and effective testing for sludge and PW samples can help to detect and then identify the reason behind increasing in the amounts of contaminants in the disposals.

Using quality control tools, such as control charts for monitoring the performance of the selected technology and other processes over their operation and production time is highly recommended.

Root cause analysis techniques are very important to identify the hidden causes of problems and that will help develop problem prevention plan and control plan.

Providing training in advanced quality design and control tools and explaining the importance of using six sigma methodology is also recommended to improve processes and systems and reduce waste during operation and production activities.

6. References

1. *U.S.Energy Information Administration*. Independent Statistics & Energy 2010; Available from: http://www.eia.gov/countries/cab.cfm?fips=IZ.
2. SOC, *Produced Water Treatment for Zubair field in SOC*, 2012, South Oil Company: Basrah, Iraq.
3. Furuholt, E., *Environmental Effects of Discharge and Reinjection of Produced Water.* Environmental Sience Research, 1995. **52**: p. 275-288.
4. Sirilumpen, M. and J.C. Meyer, *Water Reinjection for Disposal in Erawan Field*, in *SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production*2002, Copyright 2002, Society of Petroleum Engineers Inc.: Kuala Lumpur, Malaysia.
5. Navarro, W., *Produced Water Reinjection in Mature Field With High Water Cut*, in *Latin American & Caribbean Petroleum Engineering Conference*2007, Society of Petroleum Engineers: Buenos Aires, Argentina.
6. Surace, D., et al., *Installation of a Produced Water Treatment System on Raml Field (Western Desert), Egypt*, in *SPE Annual Technical Conference and Exhibition*2010, Society of Petroleum Engineers: Florence, Italy.
7. Igunnu, E.T. and G.Z. Chen, *Produced water treatment technologies.* Int. J. Low-Carbon Tech., 2012: p. cts049.

8. Evans, J.R. and W.M. Lindsay, *Managing for Quality and Performance Excellence*, in *Managing for Quality and Performance Excellence B2 - Managing for Quality and Performance Excellence*2011, Cengage Learning.
9. Evans, J. and W. Lindsay, *An introduction to Six Sigma and process improvement*, in *An introduction to Six Sigma and process improvement B2 - An introduction to Six Sigma and process improvement*2005, South-Western: Mason, OH.
10. Subber, A.R.H., M.A. Ali, and T.M. Salman, *Radon Concentration in oily Sludge Produced from Oil Refineries in the Southern oil plant at Basra Governorate-Iraq.* Archives of Applied Science Research, 2011. **3**(6): p. 263-271.
11. Cross, F.T., *Indoor radon and lung cancer, reality or myth? : Twenty-ninth Hanford Symposium on Health and the Environment, October 15-19, 1990 / edited by Fredrick T. Cross ; sponsored by the United States Department of Energy and Battelle, Pacific Northwest Laboratories*. 1992: Columbus : Battelle Press, 1992.

Design of Completely Dynamic X-Bar Process Control Procedure – Part II: Design Procedure

Nabin Sapkota[1], Hani Aburas[2] and Ahmad K. Elshennawy[1]
[1]University of Central Florida, [2]King AbdulAziz University
Nabin.Sapkota@ucf.edu

Abstract

This paper introduces a completely dynamic $\overline{X}$ process control procedure that uses alternatively two different sets of parameters. The first set considers the longest sampling interval length, the smallest sample size, and the widest control limit coefficient, while the second set includes the shortest sampling interval length, the largest sample size, and the narrowest control coefficient. The motivation behind proposing this procedure is faster detection of sources of out-of-control by making all three parameters (sampling interval length, sample size and the control limit coefficient) adaptive. Additionally, in some industrial setups, it is logical and practical not to maintain fixed design parameters owing to the natural inherent randomness in the production processes from resources utilization and effectiveness point of view.

1. Introduction

Readers are strongly advised to refer to the part I of this paper by the same authors [14]. Control chart is an important statistical process control (SPC) technique for monitoring process performance and eliminating the need for mass inspection while investigating possible operational problems thereby ensuring better process control and resource utilization. As a modification to the basic control chart, warning limits are usually added to alert users that the process might be going out of control. Traditionally, the design and operation of control charts requires the determination of three parameters: the sampling interval length (h), the sample size (n), and the control limit coefficient (k). Further enhancement to the traditional control chart can be achieved by using runs rules. In the past, authors have presented their designs of control charts with either variable sampling interval scheme (e.g., Raynolds, et al. [1], Renolds [2], [3], and [4], Reynolds and Arnold [5], Runger and Pignatiello [6], Runger and Montgomery [7], Amin and Miller [8], Renolds, Arnold and Baik [9], etc.), or control charts with variable sample interval scheme with run rules (e.g., Amin and Letsinger [10], Amin and Hemashinha [11], and Cui and Reynolds [12], etc.) or with variable sample size schemes as in Prabhu [13] for different types of control charts. Detail literature review of some of the very relevant works can be found in the review paper by Aburas et al. [14]. All earlier research in the area of dynamic or adaptive control procedures involved either parameter h or parameter n or both. None of previous research considers using k in the mix or using combinations of h, n and k together. The major contributions of this research are as follows:

1. A dynamic $\overline{X}$ control limit procedure that uses dual control limits.

2. A completely dynamic $\overline{X}$ process control procedure that uses dual sampling interval lengths, sample sizes, and control limits.
3. A combined dynamic $\overline{X}$ process control procedure in h and k and a combined dynamic $\overline{X}$ process control procedure in n and k.
4. Graphical and statistical comparisons among the different dynamic and traditional schemes developed using average time to signal.

In the subsequent sections, principle, assumptions, properties and the performance measures of the proposed procedure is explained along with the results of the comparative study with the other procedures.

2. Completely dynamic $\overline{X}$ process control procedure

2.1 Principle and assumptions

In static control charts, the three design parameters, (h, n, and k) are fixed throughout the duration of the monitoring process, regardless of the status of the process at each sampling point. Proposed completely dynamic $\overline{X}$ process control procedure allows parameters, (h, n, and k) to change in real time, taking into consideration the current sample information (Figure 1). In this procedure, if the prior observation indicates 'in-control', samples with smaller sample size are taken less frequently and relaxed control limits are adapted. On the other hand, if the previous sample static indicates a potential out-of-control condition, frequent sampling is carried out with larger sample size along with tighter control limits. The real time variation of the parameters provides opportunity to quickly detect the assignable cause shortly after it occurs, thereby, reducing the number of defective parts produced during the out-of-control period.

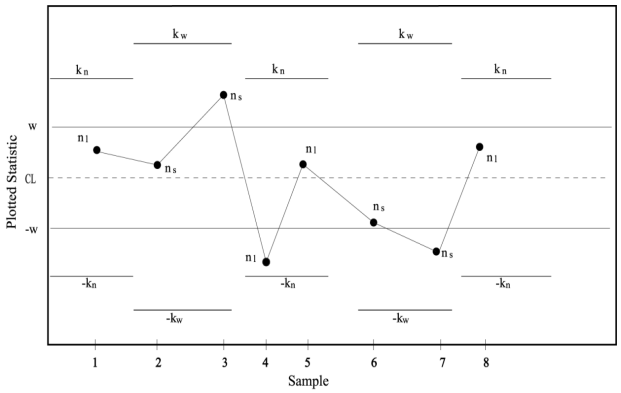


Figure 1. Proposed completely dynamic $\overline{X}$ process control procedure

In this dynamic procedure, the quality characteristic of interest is assumed to be independent and normally distributed random variable with known finite mean (μ) and variance (σ^2). Thus, the standardized variable of the i^{th} random sample, $Z_i = \dfrac{\overline{X}_i - \mu}{\sigma/\sqrt{n_i}}$, is approximately normal with mean = 0 and standard deviation = 1, where $\overline{X}_i$ and n_i are the corresponding sample mean and sample size, respectively for the i^{th} sample.

It is assumed in this procedure that the shift in the process mean, if it exists, occurs at the beginning when the process is just starting, and the process maintains the same level of shift until discovered and eliminated. After each sampling, corresponding standardized value is plotted on the control chart, and the decision is made about the status of the process based on the location of the plotted statistic with respect to the considered adapted control limits and relative to the switching limits. Besides, the size of the shift in the process mean is expressed in terms of process standard deviation units give as $\delta = \dfrac{\mu_1 - \mu}{\sigma}$.

2.2 Notations and symbols

The following notations are used to facilitate the development of the proposed procedure:

k_f = fixed control limit coefficient.
k_w = widest control limit coefficient.
k_n = narrowest control limit coefficient.
w_h = switching parameter used in a variable sampling interval scheme.
w_n = switching parameter used in a variable sample size scheme.
w_k = switching parameter used in a variable control limit procedure.
h_f = fixed sampling interval length.
h_s = shortest sampling interval length.
h_l = longest sampling interval length.
n_f = fixed sample size.
n_s = smallest sample size.
n_l = largest sample size.
v_1 = widest control limit.
v_2 = narrowest control limit.
b_1 = probability of considering the widest control limit as the start-up version.
b_2 = probability of considering the narrowest control limit as the start- up version.
k_i = control limit coefficient associated with sampling point (i).
h_i = sampling interval length associated with sampling point (i).
n_i = sample size associated with sampling point (i).
S_i = set of parameters associated with sampling point (i).

In a variable sampling interval (VSI) scheme, the sampling interval length is allowed to change between two values; the longest (h_l) and the shortest interval lengths (h_s), respectively. Assuming the process is in-control, the subsequent interval length for sample i (h_i), is determined based on the following rule:

$$h_i = \begin{cases} h_s & w_h < Z_{i-1} < k_f \\ h_l & -w_h < Z_{i-1} < w_h \\ h_s & -k_f < Z_{i-1} < -w_h . \end{cases} \quad (1)$$

Similarly, in a VSI scheme with dual sample sizes, the largest sample size (n_l) and the smallest sample size (n_s), are considered. The sample size for i^{th} sample (n_i) is based on the following rule:

$$n_i = \begin{cases} n_l & w_n < Z_{i-1} < k_f \\ n_s & -w_n < Z_{i-1} < w_n \\ n_l & -k_f < Z_{i-1} < -w_n . \end{cases} \quad (2)$$

The control limit coefficient, k_i, varies between two different values, the widest (k_w) and the narrowest control limit coefficient (k_n). At each sampling point, the next value of k_i is based on the location of the computed statistic with respect to the considered control limits and relative to the switching or threshold limits. The rule for k_i is as follows:

$$k_i = \begin{cases} k_n & w_k < Z_{i-1} < k_w \text{ or } w_k < Z_{i-1} < k_n \\ k_w & -w_k < Z_{i-1} < w_k \\ k_n & -k_w < Z_{i-1} < -w_k \text{ or } -k_n < Z_{i-1} < -w_k . \end{cases} \quad (3)$$

To simplify the implementation of the proposed completely dynamic $\overline{X}$ procedure, a single switching parameter, w, can be selected by assuming that $w_n = w_h = w_k = w$. Three rules mentioned above can be combined and rewritten as follows:

$$S_i = \begin{cases} S_2 = h_s, n_1, k_n & w < Z_{i-1} < k_w \text{ or} \\ & w < Z_{i-1} < k_n \\ S_1 = h_1, n_s, k_w & -w < Z_{i-1} < w \\ S_2 = h_s, n_1, k_n & -k_w < Z_{i-1} < -w \text{ or} \\ & -k_n < Z_{i-1} < -w. \end{cases} \quad (4)$$

As per this rule the area between the considered control limits is divided into two regions, corresponding to two different sets of parameters. The first set (S_1) includes the longest interval length, the smallest sample size, and the widest control limit coefficient where as the second set (S_2) includes the shortest interval length, the largest sample size, and the narrowest control limit coefficient, respectively.

In order to compare the performance of the proposed procedure to its traditional counterpart it is logical to assume that both procedures should perform and exhibit same characteristics in terms of expected average sampling interval length, expected average

sample size, and expected control limit coefficient. This condition can be achieved by designing the proposed completely dynamic $\overline{X}$ procedure to have parameters same as of traditional procedure (fixed interval length, fixed sample size, and fixed control limit coefficient) ensuring similar resource utilization when the process is on target (i.e., $\mu = \mu_0$). In doing this kind of parameters match-up for completely dynamic procedure with that of its traditional counterpart, the following three additional restrictions have to be considered

$$E[h_i \mid \delta = 0] = h_f. \quad (5)$$

$$E[n_i \mid \delta = 0] = n_f. \quad (6)$$

$$E[k_i \mid \delta = 0] = k_f. \quad (7)$$

The completely dynamic $\overline{X}$ procedure has eight parameters (h_s, h_l, n_s, n_l, k_n, k_w, w, and b_1) and three constraints (equations 5 through 7).

In this paper, we will use a pair of control limit coefficients to compute the switching parameter value, w. Using the third constraint, an expression to calculate the switching parameter is given by:

$$w_k = \phi^{-1}\left(\frac{2\phi(k_w)[A_k b_2 - B_k] + 2\phi(k_n)[C_k - A_k b_2] + 4B_k\phi(k_w)\phi(k_n) - C_k}{4\phi(k_n)[A_k - A_k b_2] + 4A_k b_2\phi(k_w) - 2A_k} \right), \quad (8)$$

where $A_k = k_w - k_n$, $B_k = k_f - k_n$, and $C_k = k_w - k_f$.

Applying the same approach, and based on the first and second constraints, the following expressions can also be driven to compute the switching parameter w_n:

$$w_n = \phi^{-1}\left(\frac{2\phi(k_w)[A_n b_2 - B_n] + 2\phi(k_n)[C_n - A_n b_2] + 4B_n\phi(k_w)\phi(k_n) - C_n}{4\phi(k_n)[A_n - A_n b_2] + 4A_n b_2\phi(k_w) - 2A_n} \right), \quad (9)$$

where $A_n = n_s - n_l$, $B_n = n_f - n_l$, and $C_n = n_s - n_f$.

$$w_h = \phi^{-1}\left(\frac{2\phi(k_w)[A_h b_2 - B_h] + 2\phi(k_n)[C_h - A_h b_2] + 4B_h\phi(k_w)\phi(k_n) - C_h}{4\phi(k_n)[A_h - A_h b_2] + 4A_h b_2\phi(k_w) - 2A_h} \right), \quad (10)$$

where $A_h = h_l - h_s$, $B_h = h_f - h_s$, and $C_h = h_l - h_f$.

The subscript associated with each switching parameter in Equations 8 through 10 refers to the constraint considered to derive the expression, i.e., w_h means that the switching parameter expression is driven based on the sampling interval length constraint.

Assuming that the start-up probabilities are known and starting with a pair of control limit coefficients, we are left with two more choices: the first choice is either to set the shortest interval length or the longest interval length, and the second is either to set the smallest sample size or the largest sample size. We have chosen to set the shortest interval length and the smallest sample size.

With a pair of control limit coefficients, switching parameter value, fixed and smallest sample sizes, the largest sample size can be computed by equating Equation (8) and Equation (9) and solving for n_l as shown in Equation (11).

$$n_l = \frac{k_n B n_s + k_f C n_s + n_s A k_n - n_f C k_w + n_f C k_n}{k_w A + k_f C + B k_w}. \quad (11)$$

Similarly, with a pair of control limit coefficients, switching parameter value, fixed and shortest sampling interval lengths, the longest sampling interval length can be computed by equating Equations (8) and (10) and solving for h_l (shown as Equation #12)

$$h_l = \frac{k_w A h_s + k_f C h_s + h_s B k_w - h_f C k_w + h_f C k_n}{k_n B + k_f C + A k_n}, \quad (12)$$

where,

$A = 2b_2\phi(k_w) + 2\phi(k_n) - 2b_2\phi(k_n) - 1.$

$B = 2\phi(k_w) - 2b_2\phi(k_w) + 2b_2\phi(k_n) - 4\,\phi(k_w)\,\phi(k_n).$

$C = 4\,\phi(k_w)\,\phi(k_n) - 2\phi(k_w) - 2\phi(k_n) + 1.$

2.3 Performance measures of the completely dynamic $\overline{X}$ procedure

The efficacy of a control procedure depends on how fast it identifies an out-of-control condition and signals. In traditional control procedures, the average time to signal (ATS), can be easily calculated from the ARL by multiplying it by the fixed interval length. In this proposed procedure, interval length takes more than one value thereby making this direct relationship between the ARL and the ATS invalid and demands both measures to be considered separately. For this purpose, the Markov chain approach can be applied to compute both measures. First, six different states of the system are defined as follows:

State 1: state when the standardized plotted statistic falls within the switching limits, $(w_k, -w_k)$, given that the widest version of the procedure is considered.

State 2: state when the standardized plotted statistic falls outside the switching region, but within the in-control area, given that the widest version of the procedure is considered.

State 3: state when the standardized plotted statistic falls outside the control limits, given that the widest version of the procedure is considered. This state is called an absorbing state since it cannot be exited without stopping, searching, and eliminating the assignable cause, if it exists.

State 4: state when the standardized plotted statistic falls within the switching limits, $(w_k, -w_k)$, given that the narrowest version of the procedure is considered.

State 5: state when the standardized plotted statistic falls outside the switching region, but within the in-control area, given that the narrowest version of the procedure is considered.

State 6: state when the standardized plotted statistic falls outside the control limits, given that the narrowest version of the procedure is considered. This state is also an absorbing state as in State 3.

The one step transition probability matrix P^{δ} can be written as follows:

$$\begin{bmatrix} p^{\delta}_{11} & p^{\delta}_{12} & p^{\delta}_{13} & p^{\delta}_{14} & p^{\delta}_{15} & p^{\delta}_{16} \\ p^{\delta}_{21} & p^{\delta}_{22} & p^{\delta}_{23} & p^{\delta}_{24} & p^{\delta}_{25} & p^{\delta}_{26} \\ p^{\delta}_{31} & p^{\delta}_{32} & p^{\delta}_{33} & p^{\delta}_{34} & p^{\delta}_{35} & p^{\delta}_{36} \\ p^{\delta}_{41} & p^{\delta}_{42} & p^{\delta}_{43} & p^{\delta}_{44} & p^{\delta}_{45} & p^{\delta}_{46} \\ p^{\delta}_{51} & p^{\delta}_{52} & p^{\delta}_{53} & p^{\delta}_{54} & p^{\delta}_{55} & p^{\delta}_{56} \\ p^{\delta}_{61} & p^{\delta}_{62} & p^{\delta}_{63} & p^{\delta}_{64} & p^{\delta}_{65} & p^{\delta}_{66} \end{bmatrix} \quad (13)$$

where p^{δ}_{jk} is the transition probability from state *j* to *k*. These one step transition probabilities can be calculated as follows:

$p^{\delta}_{11} = p^{\delta}_{41} = pr[-w<Z_{i-1}<w|v_1]$

$= \phi(w-\delta\sqrt{n_s}) - \phi(-w-\delta\sqrt{n_s})$.

1. $p^{\delta}_{12} = p^{\delta}_{42} = pr[w< Z_{i-1}< k_w |v_1] + pr[-k_w<Z_{i1}< -w|v_1] = \phi(k_w-\delta\sqrt{n_s}) - \phi(w-\delta\sqrt{n_s}) + \phi(-w-\delta\sqrt{n_s}) - \phi(-k_w-\delta\sqrt{n_s})$.
2. $p^{\delta}_{13} = p^{\delta}_{43} = 1- (p^{\delta}_{11} + p^{\delta}_{12})$.
3. $p^{\delta}_{24}=p^{\delta}_{54} = pr[-w<Z_{i-1}<w \mid v_2] = \phi(w-\delta\sqrt{n_1}) - \phi(-w-\delta\sqrt{n_1})$.
4. $p^{\delta}_{25}=p^{\delta}_{55}=pr[w<Z_{i-1}< k_n |v_2]+pr[-k_n< Z_{i-1}<-w |v_2]=\phi(k_n-\delta\sqrt{n_1}) - \phi(w-\delta\sqrt{n_1}) + \phi(-w-\delta\sqrt{n_1}) - \phi(-k_n-\delta\sqrt{n_1})$.
5. $p^{\delta}_{26} = p^{\delta}_{56} = 1- (p^{\delta}_{24} + p^{\delta}_{25})$.
6. $p^{\delta}_{33}= p^{\delta}_{66}= 1$.
7. Rest of the one-step transition probabilities are zero.

Applying the Markov chain approach, considering transition probabilities among the non-absorbing states (Q_δ), the average number of samples needed to detect an out-of-control condition, or ARL, is:

$$ARL_\delta = \mathbf{i}\,(\mathbf{I} - Q_\delta)^{-1}\,\mathbf{1} \qquad (14)$$

where **i** is the (1×4) row vector of the initial probabilities of being in each one of the non-absorbing states, and **I** is 4×4 identity matrix. Since $\mathbf{i}\,(\mathbf{I} - Q_\delta)^{-1}$ represents the total number of visits made to each one of the four states defined earlier, then multiplying the number of visits made to each state by its corresponding sampling interval length will yield the ATS, which can be mathematically written as follows:

$$ATS_\delta = \mathbf{i}\,(I - Q_\delta)^{-1}\,\mathbf{h}, \qquad (15)$$

where $h = \begin{bmatrix} h_l \\ h_s \\ h_l \\ h_s \end{bmatrix}$.

2.4 Properties of the completely dynamic $\overline{X}$ procedure

Applying the process control procedure as a renewal reward process mentioned in Ross, [15], the expected control limit coefficient is the ratio of the expected control limit coefficients considered to the expected number of subgroups observed before the control scheme signal. Hence,

$$E(k_i) = \frac{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\,\mathbf{k}}{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\mathbf{1}} \qquad (16)$$

Similarly, the expected sample size and the expected sampling interval length can be computed as follows:

$$E(n_i) = \frac{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\,\mathbf{n}}{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\mathbf{1}} \qquad (17)$$

and,

$$E(h_i) = \frac{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\,\mathbf{h}}{\mathbf{i}\,Q_\delta (I - Q_\delta)^{-1}\mathbf{1}} \qquad (18)$$

where the column vector **k** represents the widest and narrowest control limit coefficients; the column vector **n** represents the largest and smallest sample sizes; and the column vector **h** represents the longest and shortest interval lengths.

2.4.1 Special case I, combined dynamic $\overline{X}$ procedure in interval length and control limit:

In the proposed control procedure, parameters, k and h change in real time by considering the current sample information. Regardless of the start-up version of the procedure, the area between the considered control limits is divided into two sub-areas corresponding to two different sets of parameters. While keeping the sample size fixed, the first set includes the longest interval length and the widest control limit coefficient; the second set includes the shortest interval length and the narrowest control limit coefficient.

At the beginning of the process, and given that the control limit coefficients have been selected, the switching parameter, w, can be computed according to Equation (8). Selecting the fixed and the shortest sampling

interval lengths, the fixed, the narrowest, and the widest control limit coefficients, and also assuming that $w_k = w_h = w$, Equation (12) can be used to calculate the longest sampling interval length.

With the exception that the expected sample size will be equal to the fixed sample size since it is kept fixed and unchanged, the expected control limit coefficient and the expected interval length can be computed using Equation (16) and Equation (18), respectively. In calculating the first step's initial transition probabilities, sample size n_i will be replaced by n_f, but everything else remains the same. Due to the fact that two interval lengths are considered, the average run length obtained from Equation (14) will not be enough as a performance measure; therefore, the average time to signal must be computed using Equation (15).

2.4.2 Special case II, combined dynamic $\overline{X}$ procedure in sample size and control limit:

This is a dynamic scheme with two parameters. Based on the current sample information, the procedure changes both the sample size and the control limit coefficient in real time, keeping the sampling interval length constant. The area between the control limits is also divided into two regions according to two different sets of parameters. The first set includes the smallest sample size and the widest control limit coefficient, and the second set includes the largest sample size and the narrowest control limit coefficient.

At the beginning of the process, and given that the control limit coefficients have been specified, the switching parameter, w, can be computed using Equation (8). Selecting the fixed and smallest sample sizes, the control limit coefficients, and also assuming that $w_k = w_n = w$, Equation (11) can be used to calculate the largest sample size.

Almost all properties and performance measures presented earlier will be considered in this procedure. With the exception that the expected sampling interval length will be equal to the fixed interval length, the expected control limit coefficient and the expected sample size can be computed using Equation (16) and Equation (17), respectively. In addition, in this procedure, the average run length computed by Equation (14) will be enough to measure the performance of the procedure since the direct relationship between average run length and average time to signal is valid.

3. Comparative study and results

Comparative study of the performance of the proposed procedure and the other dynamic procedures along with the classical procedures was performed on more than 3000 simulated design sets using Matlab® software.

3.1 Simulated design sets

The parameters for the experiments were selected as follows:

a. For simplicity and comparison purpose, the fixed interval length and the fixed control limit coefficient are chosen to be one hour and three standard deviations, respectively.
b. The comparison is conducted using a broad range of shifts (.25, .5, .75, 1, 1.25, and 1.5).
c. Four short sampling interval lengths selected are: 0.01 hr, 0.10 hr, 0.30 hr, and 0.5 hr.
d. The smallest sample sizes selected based on the rule: $n_s = 1, 2, \ldots, n_{f-1}$, where $n_f = 3,4,..10$.
e. Three different switching limits selected are: 0.5, 1.0, and 1.5.

Each combination of all of these parameters is called a design set which includes the shift value, fixed, shortest, and longest interval lengths, fixed, smallest, and largest sample sizes, and fixed, narrowest, and widest control limit coefficients (δ, h_f, h_s, h_l, n_f, n_s, n_l, k_f, k_n, and k_w).

3.2 Results

For switching limits of ± 0.5, it is found that the proposed completely dynamic procedure, as shown in Figures 2 outperforms all other procedures. This is also true for other

shifts as long as $\delta \leq 0.75$. The legends used in this figure and subsequent similar figures are: C for classical, CD for completely dynamic, K for dynamic in k, NK for dynamic in n & k, HK for dynamic in h & k, and NH for dynamic in n & h.

As shift approaches 1, the combined dynamic in interval length and control limit takes the lead in terms of better performance, followed by the completely dynamic procedure. The superiority of the combined dynamic in interval length and control limit becomes dominant as shift increases, and both the completely dynamic and combined dynamic in sample size and interval length show performance similarity with a little advantage to the latter (refer Figures 3 for shift =1.25). Likewise, for switching limits of ±1.0, the proposed completely dynamic procedure, still outperforms all other procedures for $\delta \leq 0.75$.

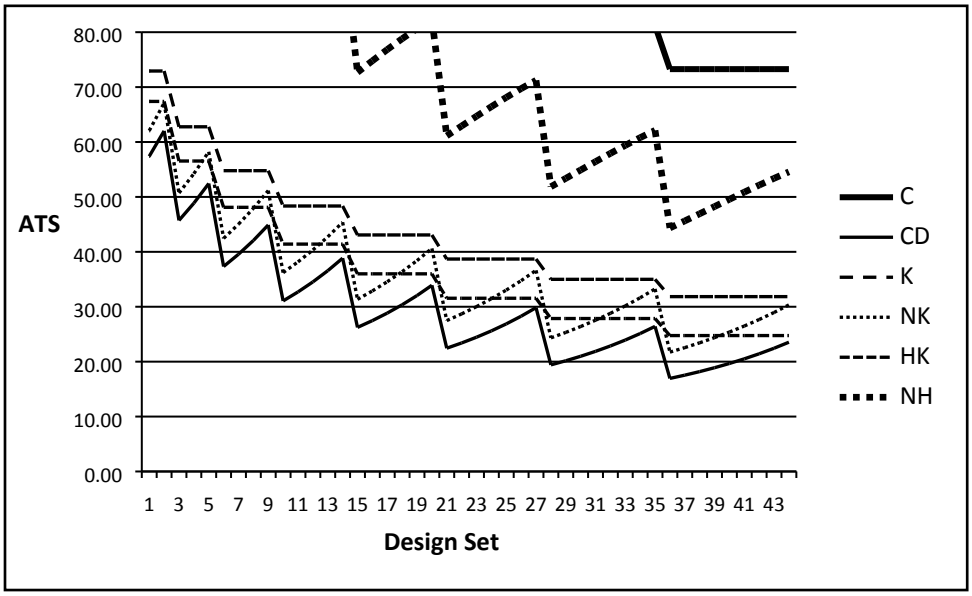


Figure 2. ATS of dynamic & classical procedures (h_s = 0.10 hr, δ = 0.25, w = ± 0.5)

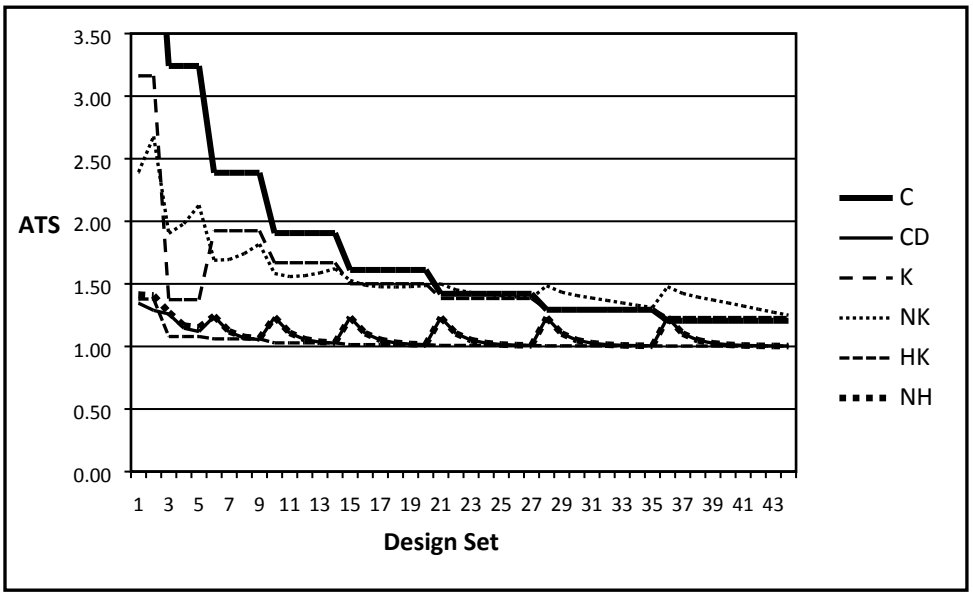


Figure 3. ATS of dynamic & classical procedures (h_s = 0.10 hr, δ = 1.25, w = $\pm$ 0.5)

As shift approaches 1, the combined dynamic in interval length and control limit shows performance dominance, and the completely dynamic continues to be the second best. For higher shifts, the combined dynamic in interval length and control limit continues to be the best in terms of faster detection of out-of-control conditions (see Figure 4).

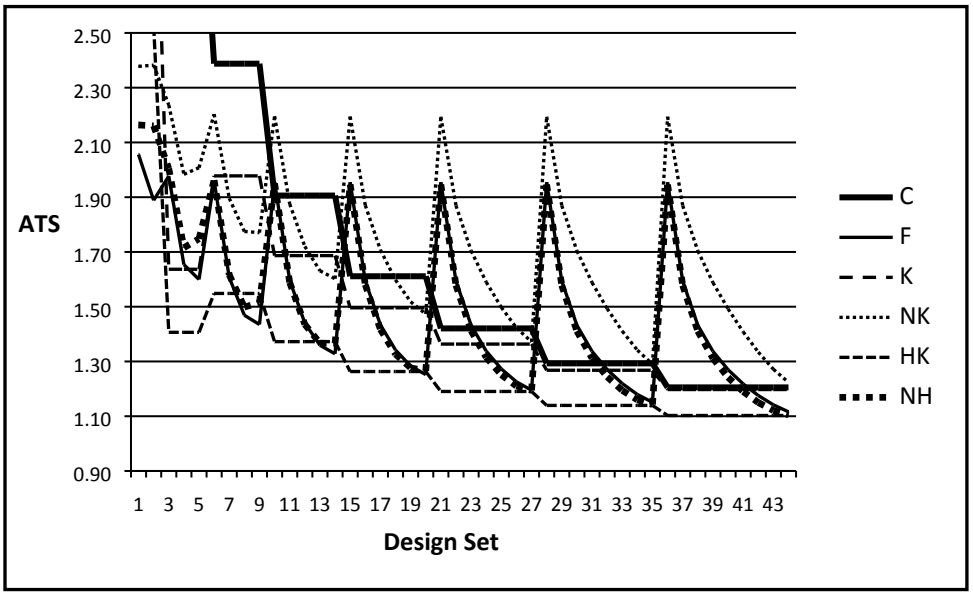


Figure 4. ATS of dynamic & classical procedures (h_s = 0.50 hr, δ = 1.25, w = $\pm$ 1.0)

Similarly as in the cases of switching limits, 0.5 and 1.0, for higher switching limit (1.5), the completely dynamic procedure continues to show superiority for small shifts ($\delta \leq 0.75$) as well. However, for higher shifts, the combined dynamic in sampling interval and control limit takes the lead in terms of better performance; see Figures 5 and 6).

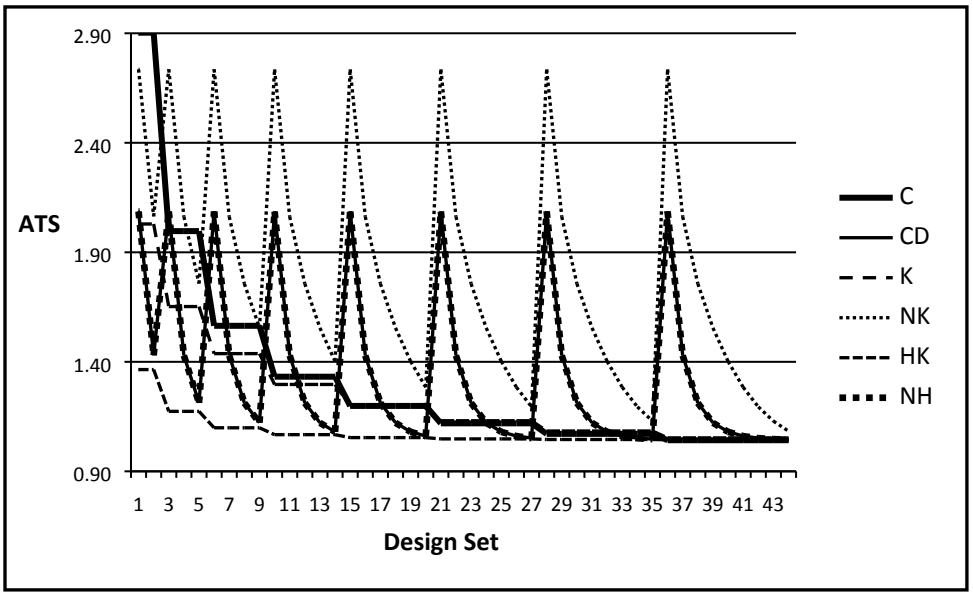


Figure 5. ATS of dynamic & classical procedures (h_s = 0.10 hr, δ = 1.50, w = $\pm$ 1.5)

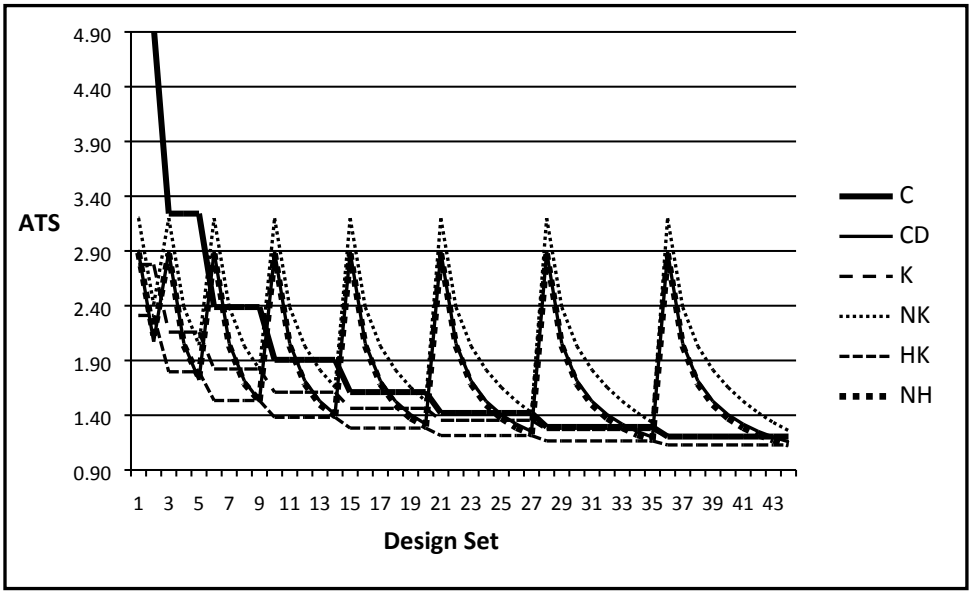


Figure 6. ATS of dynamic & classical procedures (h_s = 0.05 hr, δ = 1.25, w = $\pm$ 1.5)

As a summary, the completely dynamic process control procedure outperforms all others for a wide range of small shifts ($\delta < 1$). As shift approaches 1, the completely dynamic starts to lose its superiority to the combined dynamic in interval length and control limit, however, continues to remain as the second best. In the cases of higher shifts, the combined dynamic in interval length and control limit shows performance dominance over all other procedures.

Statistical tests of hypotheses are conducted to compare the different proposed procedures using the t-test using Minitab. A summary of the statistical results and their interpretations are shown in the Tables 1, 2, and

3. All statistical tests of hypotheses are conducted at significance level of 0.05.

Table 1. Summary of the Tested Hypotheses for Shifts 0.25, 0.5, and 0.75

Hypothesis	p-value	Conclusion
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{Shewhart procedure}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{Shewhart procedure}}$	0.00	The completely dynamic procedure is the best.
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k}}$	0.00	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k \& n}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k \& n}}$	0.00	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k \& h}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k \& h}}$	0.00	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in n \& h}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in n \& h}}$	0.00	

Table 2. Summary of the Tested Hypotheses for Shift = 1.00, 1.25, and 1.5

Hypothesis	p-value shift =1.0	p-value shift =1.25	p-value shift =1.5	Conclusion
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{Shewhart procedure}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{Shewhart procedure}}$	0.00	0.00	0.02	The dynamic procedure in h and k is the best. *
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k}}$	0.00	0.00	0.60	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k \& n}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k \& n}}$	0.00	0.00	0.00	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in k \& h}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in k \& h}}$	1.00*	1.00*	1.00*	
H_0: $\mu_{\text{completely dynamic procedure}} = \mu_{\text{dynamic procedure in n \& h}}$ H_1: $\mu_{\text{completely dynamic procedure}} < \mu_{\text{dynamic procedure in n \& h}}$	0.007	0.35	0.64	

The statistical hypothesis tested confirmed the conclusions made based on the previous graphical comparisons.

4. Discussion/Conclusion

In this paper, we introduce a completely dynamic $\overline{X}$ process control procedure. The procedure uses two different sets of parameters alternatively. The first set considers the longest sampling interval length, the smallest sample size, and the widest control limit coefficient, while the second set includes the shortest sampling interval length, the largest sample size, and the narrowest control limit coefficient. The first special case is a combined dynamic $\overline{X}$ process control procedure in interval length and control limit. It varies the control limit and interval length, but keeps the sample

size fixed. The second special case is a combined dynamic $\overline{X}$ process control procedure in sample size and control limit. This procedure varies the sample size and control limit, while keeping the interval length constant.

Graphical comparisons and statistical tests of hypotheses show that the completely dynamic $\overline{X}$ procedure outperforms others for small shifts (δ <1). However, for higher shifts, the combined dynamic $\overline{X}$ procedure in interval length and control limit shows superiority over all other procedures. Therefore, authors would like to recommend using procedure proposed in this study unless δ >1 in which case, it is logical to use control procedure dynamic in h and k.

6. Future work

In the literature, many economic designs are found for static control charts. Mostly, these economic design models have not been developed systematically resulting in various approaches and assumptions in designs. In some designs production continues during search and repair, whereas in other designs the process is shut down during repair while in some other designs the process is shut down during both search and repair. Likewise, some maximize income and some minimize cost. Authors of this paper would like to propose a new economic design for the procedure proposed in this work as a continuation of this work in future where logical and plausible assumptions will be made to arrive at the economic design for the proposed control scheme.

6. References

[1] M. R. Reynolds Jr., R. W. Amin, J. C.Arnold, and J. A. Nachlas, "$\overline{X}$ Charts with Variable Sampling Intervals", Technometrics, 1988, 30, pp. 181-399.

[2] M. R. Reynolds Jr., "Evaluating Properties of Variable Sampling Interval Control Charts", Sequential Analysis, 1995, 14, pp. 59-97.

[3] M. R. Reynolds Jr., "Shewhart and EWMA Variable Sampling Interval Control Charts with Sampling at Fixed Times", Journal of Quality Technology, 1996, 28, pp. 199-212.

[4] M. R. Reynolds Jr., "Variable- Sampling-Interval Control Charts with Sampling at Fixed Times", IIE Transactions, 1996, 28, pp. 497-510.

[5] M. R. Reynolds Jr., and J. C. Arnold, "Optimal One-Sided Shewhart Control Charts with Variable Sampling Intervals", Sequential Analysis, 1989, 8, pp. 51-77.

[6] G. C. Runger, and J. J. Pignatiello Jr., "Adaptive Sampling for Process Control", Journal of Quality Technology, 1991 23, pp. 135-155.

[7] G. C. Runger, and D. C. Montgomery, "Adaptive Sampling Enhancements for Shewhart Control Charts", IIE Transactions, 1993, 25, pp. 41-51.

[8] R. W. Amin, and R. W. Miller, "A Robustness Study of $\overline{X}$ Charts with Variable Sampling Intervals", Journal of Quality Technology, 1993, **25**, pp. 36-44.

[9] M. R. Reynolds Jr., J. C. Arnold, and J. W. Baik, J. W., "Variable Sampling Interval $\overline{X}$ Charts in the Presence of Correlation", Journal of Quality Technology, 1996, 28, pp. 12-30.

[10] R. W. Amin, and W. Letsinger, "Improved Switching Rules in Control Procedures Using Variable Sampling Intervals", Communications in Statistics-Simulation and Computation, 1991, 20, pp. 205-230.

[11] R. W. Amin, and R. Hemasinha, "The Switching Behavior of $\overline{X}$ Charts with Variable Sampling Intervals", Communications in Statistics-Theory and Methods, 1993, 22, pp. 2081-2102.

[12] R. Q. Cui, and M. R. Reynolds Jr., "$\overline{X}$ Charts with Runs Rules and Variable Sampling Intervals", Communications in Statistics-Simulation and Computation, 1988, 17, pp. 1073-1093.

[13] S. S. Prabhu, D. C. Montgomery, and G. C. Runger, "A Combined Adaptive Sample Size and Sampling Interval $\overline{X}$ Control Scheme." Journal of Quality Technology, 1994, 26, pp. 164-176.

[14] H. M. Aburas, N. Sapkota, and A. Elshennawy, "Design of Completely Dynamic X-Bar Process Control Procedure-Part I: A Review", The Journal of Management and Engineering Integration, 2013, 6 (2). (Paper accepted December 2013).

[15] S. M. Ross, "Applied Probability Models with Optimization Applications", 1970, Holden-Day, Inc.

Developing Mobile Apps

Mark Smith
Purdue University North Central
mlsmith@pnc.edu

Abstract

This paper provides an overview of the mobile app market and a quick look at what it takes to develop a mobile application. After a review of the platform architecture and application life cycle, a quick review of the Eclipse integrated development environment will explore the major components required to develop an Android 4 application. The paper looks at functionality on the mobile device, with a focus on events for Activities and Services.

1. Introduction

The rapid growth of mobile applications and the interest in this technology by businesses and other organizations have made it imperative that Information Technology curricula include courses on mobile applications. This project began as an investigation of what environment to use to implement a Mobile Apps Development course into the current computer technology program at Purdue University North Central.

Apple and Google opened their respective mobile app stores in 2008 and 2009, and since that time over one and a half million applications have been developed for their combined platforms [1]. This paper begins with an overview of the mobile app market and then presents a case for developing mobile applications in the Android environment.

2. The Mobile App Marketplace

There are five major platforms for developing mobile applications, which include Android, iOS, RIM (Blackberry), Symbian, and Windows Mobile [2]. While all of these platforms use some form of the C programming language and/or Java for development of their apps, each has its own development environment and programming interface. Table 1 shows details on each platform in the marketplace, its core language and development environment.

Given the diversity of development options, we felt it was necessary to choose a single platform if we wanted a comprehensive course in mobile application development. In terms of market share, Android and iOS are the biggest players and none of the other participants are close, as shown in Table 2 below. This made it easy to narrow the field to those two platforms. The next section discusses how the final decision was made.

Table 1 - Mobile OS Platforms

Mobile Platform	Core Language	Environment
Android	Java or C++	Eclipse
iOS (Apple)	Objective-C	Xcode
RIM (Blackberry)	Java	Eclipse
Symbian	C++	Multiple choices
Windows Mobile	C#	VisualStudio 2010

Table 2 - Mobile OS Market Share [3]

Smartphones			4Q '11	4Q '12
(mil. phones)	4Q11	4Q12	Share	Share
Android	77.1	144.7	51%	70%
iOS	35.5	43.5	24%	21%
Blackberry	13.2	7.3	9%	4%
Microsoft	2.8	6.2	2%	3%
Others	21.8	6.0	15%	2%
Total	150.2	207.7	100%	100%

3. Choosing a Development Platform

More recent market share information shows the number of Android devices in use today dominates the market with approximately 75% share [4]. Based on this alone, Android would seem an obvious choice. However, Apple's iOS was still a serious competitor due to its early leadership position and integration capability with other Apple devices.

The software development environments for both Android and iOS are freely available and can be downloaded from the internet at no charge. This initial finding made them equally attractive for the new course. However upon further investigation, in order to make applications developed by the students generally available, Apple requires a $99 annual fee for deployment through their Apple Store. This fee also provides two instances of technical support from Apple [5]. On the other hand, open-source Android application deployment is totally free through the Google Play store and other distribution outlets. Technical support is also provided free by the open-source community.

Another consideration in determining which development platform to use was the issue of longevity. Which environment is likely to be around in the future and which environment is likely to provide the best potential for a future job? Both Android and iOS are well established at this point and have solid technical support behind them. The following figure shows that the number of Android apps surpassed the number of iOS apps in 2012.

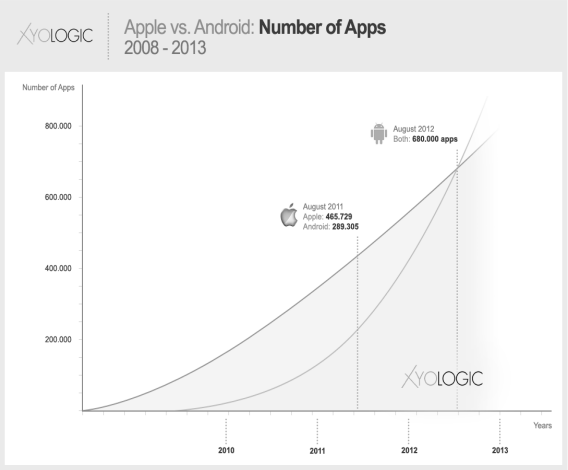


Figure 1 - Android vs. iOS Apps [6]

The last consideration for the decision was how to demonstrate the work of the students to assess their ability to develop a mobile app. Both Android and iOS have mobile device emulation as a component of their development environments, but in each case the emulator is quite slow and not able to provide all of the features available on many mobile devices.

There was also a preference for tablet devices for the course, which make available a greater viewing area for apps that businesses and other organizations are more likely to use. Since funding resources were not going to allow the university to provide devices to each student in the class, the decision was made for students (or teams of students) to provide their own devices.

In the past several months there have been many advertisements for Android tablets in the 7" display area range for around $100. This is considerably less than any Apple iOS devices that are for sale today. The purchase and testing of one of these Android devices found it suitable for the course.

All of the elements considered for the course platform pointed to Android as the most appropriate development environment for the class.

4. Android Infrastructure

Before we look at specifics for writing Android apps, it is useful to take a look at the layered infrastructure to understand how our applications will be working with the mobile device.

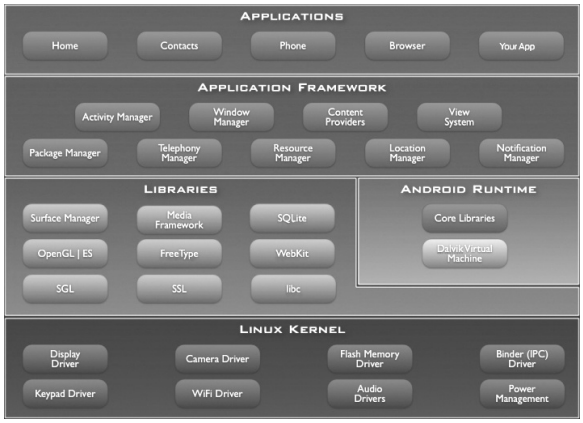


Figure 2 - Android Infrastructure [7]

The Android infrastructure starts with a Linux Kernel that provides several important features. First is a hardware abstraction layer which is a set of software routines that provide programs with direct access to hardware resources. This allows programmers to write device-independent applications. It also contains memory management and process management functions, which includes device drivers for various hardware components and networking.

The next layer contains native libraries. These include the Bionic libc, a small (and fast) derivation of the standard C library. It was developed by Google specifically for the Android embedded operating system. Next is the Surface Manager, which is used to compose windows graphics. There is also 2D and 3D graphics support/simulation, and media codecs for audio/video support. Database management support is provided by SQLite object classes and there is a WebKit library for fast HTML rendering to display web pages, as well as methods for navigation and other useful tools.

The Android runtime layer includes a set of core libraries that provide functionality available in the core libraries of Java. Each Android application runs in its own process, with its own instance of the Dalvik virtual machine (VM). The Dalvik VM uses the Linux kernel for underlying functionality such as threading and low-level memory management.

Apps are developed using tools supplied by the Application Framework, which is a toolkit developed primarily by Google, although it may also contain extensions or services provided by users. Within the framework: the Activity Manager controls the life cycle of the app, Content providers encapsulate data that is shared, the Resource Manager handles non-code resources, the Location Manager provides location services of the device (GPS, GSM, Wi-Fi), and the Notification Manager informs the user of events such as arriving messages, appointments, and other things that happen in the background.

5. Android Developer Tools Bundle

Before you begin writing Android apps, you need the appropriate software development tools. These include an integrated development environment which contains the software to code and build applications, a test environment, and the ability to deploy apps for general distribution. All of these tools can be found in the Android Software Development Kit (SDK), which can be freely downloaded from the Android Developers site supported by Google. The Android SDK provides you with the API libraries and developer tools necessary to build, test, and debug apps for Android. You can download individual software components or the ADT Bundle, which includes the essential Android SDK components and a version of the Eclipse IDE with built-in ADT (Android Developer Tools).

The ADT Bundle includes everything you need to begin developing apps on Windows, specifically:

- Eclipse + ADT plug-in
- Android SDK Tools
- Android Platform-tools
- The latest Android platform
- The latest Android system image for the emulator

You can get the ADT Bundle at:
http://developer.android.com/sdk/index.html

6. Android Application Components

Android applications are written in Java and compiled with any data and resource files into an Android package. Once installed on a device, each Android application runs in isolation from other applications, via its own Virtual Machine. Android starts the process when any of the application's components need to be executed and shuts it down when no longer needed. Each application has access only to the components that it requires to do its work, in a secure environment.

Sharing data and other system resources is managed by Android through permissions granted at the time the application is installed on the system.

6.1. Activities

An activity represents a single display with a user interface. Applications consist of one or more activities that work together to form an interconnected user experience.

The Android Activity Life Cycle contains several exit and entry points that are available to programmers as the application transitions between different states of its lifecycle (see Figure 3). These methods are necessary to allow the user to navigate through, into and out of your app.

Within these lifecycle methods, you program how your activity behaves when the user leaves and re-enters the activity. For example, you might pause music in an audio player when a user leaves your app, and then restart it when they return.

6.2. Services

A service is a component that runs in the background to perform long-running operations or to perform work for remote processes. There is no user interface with a service. Applications can initiate/communicate with services.

A service runs in the same process as the application of which it is a part and provides two main features: a method for the application to ask the system to schedule work in the background, to be run until the service or someone else explicitly stop it, and a method for an application to allow some of its functionality to be accessed by other applications.

The Android system will try to keep the process hosting a service around as long as the service has been started or has clients bound to it.

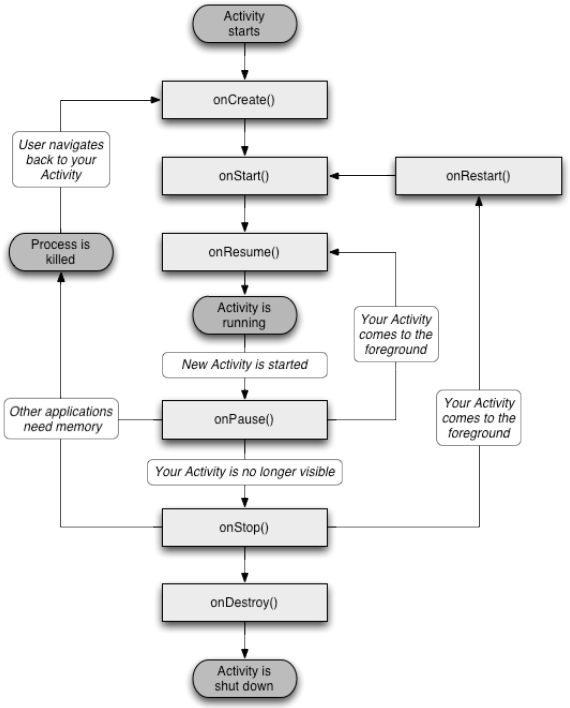


Figure 3 – Android Activity Lifecycle [8]

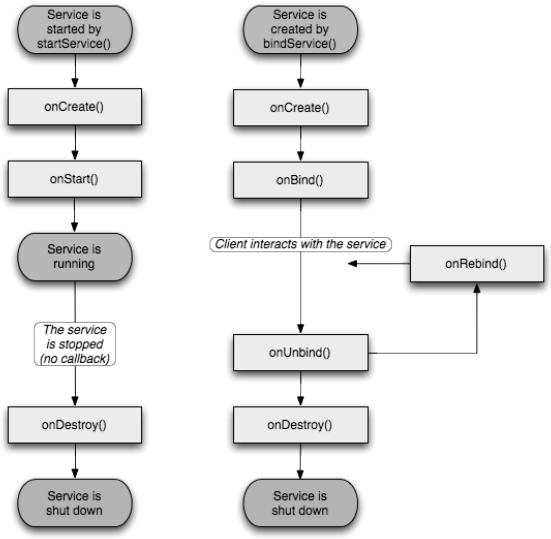


Figure 4 - Android Service Life Cycle [9]

6.3. Content Providers

Content providers manage access to data. They provide structure through encapsulation of the data, and a means of defining data security. Content providers are the standard interface between processes for sharing data.

6.4. Broadcast Receivers

A broadcast receiver is a component which allows you to register for system or application events. There are two major classes of broadcasts that can be received: Normal broadcasts, which are sent asynchronously to all receivers in no particular order, and Ordered broadcasts, which are delivered to one receiver at a time. Each receiver executes in turn, either propagating the message to the next receiver or aborting the broadcast. Android controls the sequence through a prioritizing scheme.

7. Android Connectivity

Android provides APIs to allow your apps to connect and interact with other devices over Bluetooth, NFC, Wi-Fi Direct, USB, and SIP, as well as with standard network connections.

8. Android App Development

Application development for the Purdue University North Central mobile app development course will be based on the Eclipse IDE with the ADT Bundle described earlier in this paper. This section describes some of the major components of the IDE that relate to Android app development. These components must work together to define the application, which means the programmer needs to keep them synchronized.

8.1. The Manifest

Each application has an Android Manifest file, which presents information about the

application to the Android system. The manifest provides the following information:

- A unique name (identifier) for the Java package
- The components of the app (activities, services, broadcast receivers, and content providers)
- Which processes will host application components
- Permissions
- The minimum Android API required to run the app
- The libraries needed by the app

8.2. Layouts

Layouts define the visual structure for a user interface using XML. It declares what the user sees and objects with which the user can interact. An app may have one or more layouts.

Layout files enable you to separate the presentation of your application from the code that controls its behavior, allowing you to modify them without having to change your source code and recompile. You can also create XML layouts for different screen orientations, making your apps more portable.

8.3. Java Code

Java classes are compiled into the Dalvik VM (discussed in Android Infrastructure), which was designed specifically for Android. Dalvik VM is a register-based architecture rather than stack-based Java VM. In any case, Java programmers will be very comfortable with the programming environment.

8.4. Resources

The Android resource system manages all of the non-code assets associated with an application. Application resources are compiled into the application binary at build time.

Using application resources makes it easy to update various objects of your application without modifying source code. You may also specify alternative resources to enable your application to run on different device screen sizes.

8.5. Android Virtual Device Manager

An Android Virtual Device lets you model a real device by configuring hardware and software options to be emulated by the Android Emulator. An Android Virtual Device is made up of:

- A hardware profile
- Mapping to a system image (Android platform level)
- Other options like screen dimensions, appearance, etc.
- Dedicated storage area (on your development PC)

You may create as many Android Virtual Devices as you need and use them as required to test your applications. Android Virtual Devices tend to run rather slowly, so give them time to start up and run before you decide you have a system problem.

9. Android Resources

There are many books available that cover Android application development. They are wide-ranged in terms of the audience they are geared for and the quality of coverage. There are a number of higher-education level textbooks that include student and instructor resources like sample programs, instructor manual, PowerPoint presentations, exams, and solution files [10].

Online documentation is also readily available for Android App development. In fact, much of the information is this report was

gleaned from Google's Android Developer's website, located at http://developer.android.com

10. Conclusion

For Purdue University North Central, we believe Android is the right choice for our new Mobile App Development class. Learning objectives and content for this course are still under development and the first class has been scheduled for the Spring 2014 semester. We are confident of a great educational opportunity for our students and look forward to assessing the results of this endeavor in a future paper.

11. References

[1] "Who's Winning, iOS or Android? All the Numbers, All in One Place", Technologizer By Harry McCracken, April 16, 2013, http://techland.time.com/2013/04/16/ios-vs-android/

[2] "Mobile Application Developer", March 20, 2013, http://www.itcareerfinder.com/it-careers/mobile-application-developer.html

[3] "Android Solidifies Smartphone Market Share", Forbes Tech, Chuck Jones, 2/13/2013, http://www.forbes.com/sites/chuckjones/2013/02/13/android-solidifies-smartphone-market-share/

[4] "As Android hits 75% market share, can anyone tell me why this is not Mac vs PC all over again?", VB/Mobile, John Koetsier, November 1, 2012, http://venturebeat.com/2012/11/01/as-android-grabs-75-market-share-can-anyone-tell-me-why-this-is-not-mac-vs-pc-all-over-again/

[5] "Program Enrollment", Apple Developer, Support Center, April 29, 2013, https://developer.apple.com/support/ios/enrollment.html

[6] "Xyologic: Android will overtake Apple in monthly downloads in June 2012", Ingrid Lunden, XYOLOGIC, October 11, 2011, http://paidcontent.org/2011/10/11/419-got-app-analytics-xyologic-now-has-multi-platform-numbers-for-29-countr/

[7] "How Android Works: The Big Picture", zdNet, Ed Burnette, January, 28, 2008, http://www.zdnet.com/blog/burnette/how-android-works-the-big-picture/515

[8] "Activity", Android Developers, April 29, 2013, http://developer.android.com/reference/android/app/Activity.html

[9] "Services", Android Developers, April 29, 2013, http://developer.android.com/guide/components/services.html

[10] Hoisington, Corinne, "Android Boot Camp for Developers Using JAVA Comprehensive", Course Technology, Boston, MA, 2013.

Dynamic Free Text Keystroke Biometrics System for Simultaneous Authentication and Adaptation to User's Typing Pattern

Alex King and Paulus Wahjudi
Marshall University
king184@marshall.edu, wahjudi@marshall.edu

Abstract

This paper describes a new algorithm for free text keystroke authentication and its implementation. Recent research describes an implementation of the Scaled Manhattan distance for continuous authentication. This paper expands upon that research by describing a modified method for keystroke authentication that also implements decision tree learning to dynamically adapt to changes in the authentic user's typing pattern while still detecting impostors. Additionally this new method authenticates users based on five different temporal features of key digraph events: interval, dwell time, latency, flight time, and up to up.

1. Introduction

This paper describes a new approach to developing a dynamic learning desktop application that uses free text keystroke biometrics to continuously authenticate a user once that user is logged on to system. Users are authenticated based on five different temporal features of key digraph events: interval, dwell time, latency, flight time, and up to up.

This involves two distinct phases. The first phase is the enrollment phase. In this phase, a unique template is developed to identify a particular user. The second phase is the authentication phase. During this phase, the current user's typing pattern is statistically compared to the authenticated user's template using a modified Manhattan distance metric. A user-defined confidence threshold is used to determine whether or not the current user is an imposter. If the current user is determined to be an imposter, the system will lock out that user.

Also during the authentication phase, decision tree learning is used to dynamically "learn" changes to the authentic user's typing pattern and modify the template accordingly. This way the system will not only determine whether or not the user is an imposter, but it will also continuously adapt to changes in the authenticated user's typing pattern.

This ability of the system to adapt to a user's typing pattern while simultaneously providing authentication is crucial for this system, since it is designed to be used over a long period of time. Additionally, this adapting capability allows users to begin using the system with less training than otherwise required, because the training can continue even during authentication. This will be most helpful to users who need a functioning authentication system in a shorter amount of time.

2. Background

Keystroke biometrics is the process of authenticating users based on how they type. The process begins with an enrollment phase where the system learns specific typing patterns of a user. After that, a template of the most stable or consistent patterns is created. This is used to authenticate users during the authentication phase [4]. Figure 1 (below) outlines the flow of the phases within the system.

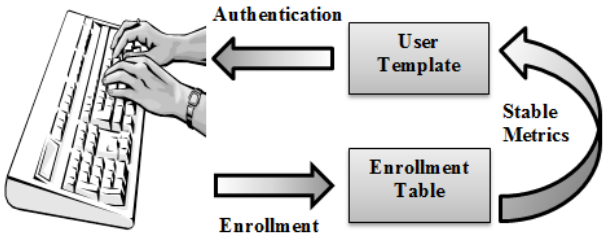


Figure 1: Keystroke Biometric Phases

The typing features analyzed in this paper relate to timing metrics between two keystrokes (or digraphs). However, it is possible to look at trigraphs (three keystrokes) or, more generally, n-graphs [2].

There are five major temporal features: interval time, dwell time, latency, flight time (down to down), and up to up [16]. These features together compose a unique typing "fingerprint" for each particular user.

In addition, there are two main categories of keystroke biometrics authentication: static (fixed text) and continuous (free text) [8]. Static or fixed text authentication provides authentication based on a specific text that the user types (such as a username or password). Alternatively, continuous (free text) authentication provides authentication continuously, no matter what text the user is actually typing [8]. This paper focuses on the continuous authentication.

This system uses a modified Manhattan distance described Bours [3]. Killourhy and Maxion found it to be one of the best performing keystroke authentication algorithms in terms of equal error rate [11]. Zhong and Dheng improved on this algorithm by combining it with the Mahalanobis distance algorithm [19].

Other research into continuous authentication has utilized a k-nearest-neighbor (*kNN*) algorithm [13][17]. In addition, more recent research describes using keystroke evaluation on mouse clicks [1][15].

2.1. Problem

Three main problems are addressed in this paper. First an algorithm needed to be developed that uses all five digraph keystroke features for continuous authentication rather than just two or three. Second, the algorithm must provide simultaneous authentication and adaptive learning so the authentication can evolve with the genuine user. Third, this algorithm must keep a running calculation of digraph metrics and their standard deviations.

3. Proposed Solution

The proposed solution will be divided into three parts addressing each of the problems listed above.

3.1. Processing Five Keystroke Features

The system uses five keystroke features as described by Shimson et al [16]. Other systems typically use only two or three keystroke features.

The following is a description of each of the keystroke features. Dwell time refers to how long a subject presses down on a particular key. Interval time refers to the time interval between when a subject lifts up on one key to when they push down on the next key. Note that metrics such as interval time can be negative, if a subject

pushes down on the second key before lifting up on the first.

Latency refers to the time between when a subject pushes down on key one to when they lift up on key two. "Up to Up" refers to the time between when a subject lifts up on key one to when the subject lifts up on key two. Similarly, flight time (or "Down to Down") refers to the time between when a user pushes down on key one to when that user pushes down on key two [16].

The following equations show how to calculate the various keystroke features:

DwellTime = KeyOneUp – KeyOneDown
IntervalTime = KeyTwoDown – KeyOneUp
Latency = KeyTwoUp – KeyOneDown
FlightTime = KeyTwoDown – KeyOneDown
UpToUp = KeyTwoUp – KeyOneUp

All five distance metrics were uses to create a scaled Manhattan distance vector similar to the one described by Bours [3].

$$= \left\{ \begin{array}{l} | \quad |/ \\ | \quad |/ \\ | \quad |/ \\ | \quad |/ \\ | \quad |/ \end{array} \right\}$$

In this case, each of the subscripts refers to a specific keystroke feature; μ refers to the stored mean in the user template; t refers to the input time from the current subject; and + refers to the stored standard deviation. Also, divide by the standard deviation because those features with the less deviation are more consistent and therefore more useful in determining an appropriate distance.

As mentioned above, only "stable" metrics can actually be used for authentication. This system determines stable metrics as those metrics where that have been recorded at least fifty times, and where the standard deviation divided by the mean is less than 0.1. This method of finding stable metrics is based off the method described by Bours and Barghouthi [4].

3.2. Running Standard Deviation

Another issue that needed to be addressed was how to keep a running standard deviation. The system does not store every single keystroke event that occurs, so it needs to determine a new standard deviation every time a new keystroke event occurs without knowing all the previous keystroke events.

To do this the system stores the mean, standard deviation, and count for each keystroke feature, and when a new keystroke event occurs the system to recalculates the standard deviation.

Devore [5] provides the following equations for computing deviation and variance:

$$[i] \qquad ^2 = \frac{(\quad)^2}{1} = \frac{\quad}{1}$$

$$[ii] \qquad = \quad ^2 \frac{(\quad)^2}{\quad}$$

Equation [i] shows how to compute the variance for a given set of data, and equation [ii] provides a "computable" formula for the numerator of the variance [5]. The above equations were used to develop the following steps (outlined in pseudo code) for computing variance.

```
If oldCount       != 0:
 newMean = ((oldCount*oldMean)+metricValue)
         /(oldCount +                  1);
 oldVariance       = oldSDev*oldSDev;
 sumXiSquared = (oldVariance*(oldCount-1))
          +(oldMean  ²* oldCount);
 sXX    = (sumXiSquared+(metricValue        ²))
          -(nMean  ²*(oldCount+1));
 newVariance      = sXX / oldCount;
 Math.Sqrt(newVariance);
```

Figure 2: Running Variance Pseudo Code

The above pseudo code shows how to compute a new standard deviation for each

keystroke event with only the previous standard deviation, previous count, and precious mean.

Zhang and Guan [18] provide a more detailed description of computing a running variance for large data streams. Also Knuth discusses and algorithm for computing variance in *The Art of Computer Programming* [12].

3.3. Simultaneous Authentication and Adaptive Learning

Another key feature to this system is that it provides simultaneous adaptive learning and authentication. This is done by using a decision tree learning mechanism. Figure 3 (below) outlines the decision tree used by the system.

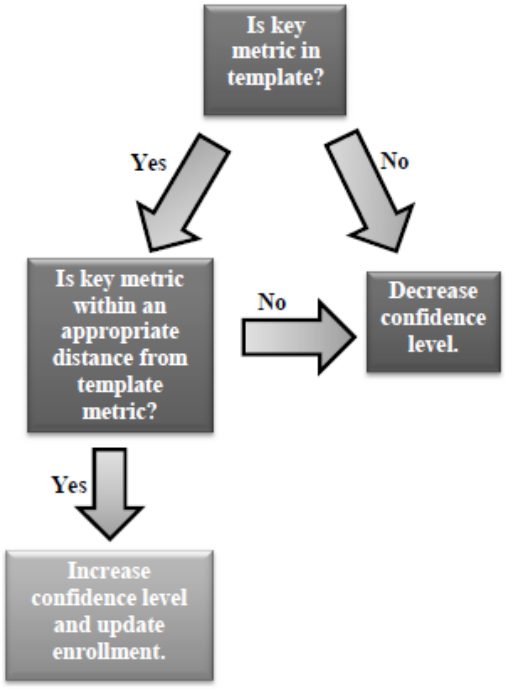


Figure 3: Decision Tree Learning

The idea behind this decision tree is that the template can be made more accurate by collecting more user data even after the authentication has begun. This essentially combines the two phases of authentication and enrollment.

Kang et al [9] propose a similar system, and they refer to it as a "Growing Window." Giot et al [6] also develop a system that "retrains" the template, and in their case, they use a different distance threshold for authentication than they do for adding it to the template.

4. Implementation and Features

This system is written in C# and designed to run on the Microsoft Windows operating system. Microsoft Visual Studio was used at development environment and SQLite is used as the system's backend database.

In terms of implementation, four major features need to be addressed. First an accurate system timer is needed, and it must not take up too much processing power. Second, the system must chronologically store and process keystroke data without violating user privacy. Third, the system must have an appropriate user interface that allowed users to tailor the system to their needs while still having a streamlined user interface. Lastly, the issue of deployment system security must to be addressed.

4.1. System Timer

One major feature of our system is that it will operate without slowing down or interrupting the genuine user's workflow. Also, it is designed so that it does not require any extra hardware beyond a typical workstation (i.e. no fingerprint scanners, token readers, etc.)

This became an issue however when finding the best way to record keystroke times. Killourhy and Maxion [9] explore this issue in detail. They found that a Windows system timer is only accurate to ~15ms, and that more accurate results could be obtained from a more precise timer [9].

Possible solutions to getting a more precise timer precision include implementing a microsecond timer, modifying the timer thread

affinity, or using an external timer. Since each of these solutions either requires extra hardware or has the potential to interrupt or slow down the users workflow, this system utilizes the Microsoft Windows 7 system timer that already runs on the Windows systems.

The timer is designed so that every so often it restarts to keep keystroke timestamps within a manageable range for the system.

4.2. Keystroke Data Processing

When processing a user's keystroke data, it is important to keep in mind that keys are not necessarily hit in consecutive order. For instance, if a user types "USA" they may hold down the shift key, and not lift up until after pushing down on the "A" key.

To account for situations such as this, the system keeps a linked list of key data objects. Each key data object contains the key's value, a "key down" time (in milliseconds), and a "key up" time. Every time a key pushed down, a key data object is created (with a placeholder for the key up time), and that key data object gets added to the linked list.

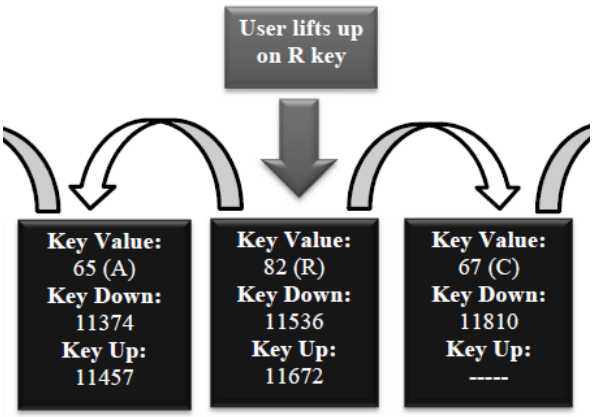


Figure 4: Authentication Example: A freeze frame of the specific moment that a subject lifts up on the 'R' key when typing 'MARCO'

Every time a user lifts up on a key, the system finds the corresponding key data object in the linked list and updates the key up time. Next the system looks to the next key data object in the linked list. If that key data object exists and it has a valid key up time (no just a placeholder), digraph metrics can be processed between the two keys

Then the system looks to the previous key data object in the list, and performs the same operation. One digraph metrics are processed on both sides of the key data object, that object is removed from the linked list. Thus no chronological record of what characters the user actually types is persistently stored.

```
user lifts up          on ThisKey       when typing
find ThisKey       in linked        list
update ThisKey UpTime
if NextKey      object exists           && has    UpTime
  process metrics between key              objects
    if NextNextKey exists &&                 has UpTime
      remove NextKey
      if PreviousKey exists &&               has   UpTime
            process    metrics between key          objects
    if PrePreviousKey exists                    &&  has   UpTime
      remove PreviousKey
      if metrics        on   both sides        of   ThisKey
processed
       remove ThisKey
```

Figure 5: Key List Pseudo Code

Also outliers are removed and to avoid recording breaks in typing, any metrics greater than ~1000ms are removed. This is another variable that could be increased or lessened based on user preferences, but at this point in time is not part of the user controls (see section 4.3).

Additionally, certain keys (such as the backspace key) are not used or identification, because their usage does not lend themselves to uniquely identifying a user.

4.3 User Controls

Another major feature of this system is a simplified user interface with a minimal number

of variables that need toggled by the user. To accomplish this, a trust threshold (described by Bours [3]) is used. Originally it was planned to be the only variable toggled by the user, but after testing, it was necessary to allow the user set a "threshold distance."

Below is a depiction of the user controls for the system. Note that the experimental system has a working name of "LockoutII." Figure 6 (below) is shows a screen shot of the user controls for this system.

Figure 6: User Controls

The threshold distance is the threshold distance from the template that the system uses to decide whether to raise or lower the confidence level. In other words, if the mean Manhattan distance is less than the threshold distance, the confidence will be raised; otherwise, the confidence will be lowered.

Even though both the threshold distance and the trust threshold relate to balancing false positive and false negative user identifications, they are both useful for test subjects to find their own balance between security and usability. The distance threshold allows the user to affect greater change on the system, while the trust threshold allows for more fine-grained control.

4.4 System Security

When developing a system used for security one needs to address how to prevent impostors from tampering with the system. If an impostor knows this system is running on a user's computer, they could easily kill the process.

One way around this would be to have the system run as an alternate data stream [14]. However, such behavior could be deemed malicious by an end user system.

Similarly, imposters can still remove the system from the end user workstation if the victim user has administrator privileges. Much like with the alternate data streams, it is not feasible to make this system so that could not be removed from the workstation once installed.

4.5 System Validation

System validation was executed by having multiple users enroll in the system. Multiple trials were staged where imposters tried to gain access. The trials were predominately performed using laptops running Windows 7. Note that future validation will require more quantification of the test trials. That way a baseline can be to compare against future releases of the system.

5. Analysis and Conclusion

This paper describes and outlines the implantation of a new method of continuous authentication that simultaneously adapts to the

genuine user's typing pattern. The proposed system effectively provides an added layer of security by ascertaining legitimate users from a pool of imposters.

The proposed system opens the door to a new generation of keystroke biometric systems by delivering several unique features not yet available in other biometric systems. This includes simultaneous adaptive learning and continuous authentication, client side operation, open source code, and authentication based on five separate keystroke features.

Additionally, the system authenticates users based on all five keystroke features described by Shimson et al [16], and the system implements a punishment and reward mechanism very similar to the one developed by Bours and Barghouthi [4].

Furthermore, the adaptive learning feature is able to conclusively mold the system to the legitimate user's typing pattern over time. Moreover, when the adaptive learning option is utilized, there is no subsequent gain in the false-acceptance or false-rejection rates in the system. Thus users can apply adaptive learning without taking any losses to security.

Even so, several issues had to be addressed in this paper. These issues included processing keystroke features, keeping a running standard deviation, performing simultaneous authentication and adaptive learning, using the Microsoft Windows 7 system timer, and minimizing user controls while still allowing for personalized settings.

Also it was found that the system works best in a stable work environment. Therefore, users with desktop workstations or users who operate consistently in the same work area will find this system more useful than those who do not work in a consistent place.

6. Future Work

With this system in particular, attaching an in-memory database for the template may enhance efficiency. Another feature that did not get added to this system is a timed lockout, where users can set a specific time for how long the system stays locked when a lockout actually occurs.

Another promising area to explore involves using mouse dynamics authentication in conjunction with keystroke authentication [1][15]. A user can be authenticated via mouse-clicks by performing a specific set of tasks using a mouse or touchpad –similar to how fixed text authentication requires the user to perform a specific set of keystrokes [15]. In this scenario, the authentication system could prompt the user to perform mouse click authentication in lieu of locking out the user.

Moreover, future research will involve refining keystroke authentication algorithms for implementation, analyzing which keys or keystroke features provide the most accurate authentication, and evaluating how typing changes during various computing tasks.

Additionally, more data needs to be generated to sufficiently test free text authentication algorithms. Specifically, this system needs to be modified so it can potentially be tested on other keystroke benchmark data sets such as the ones developed by Zack et al [17] and by Giot et al [7].

7. References

[1] Ahmed, A.A.E., Traore, I., "Anomaly intrusion detection based on biometrics," Information Assurance Workshop, 2005. IAW '05. Proceedings from the Sixth Annual IEEE SMC, 2005 pp.452,453.

[2] Bergadano, F., Gunetti, D., & Picardi, C. "User authentication through keystroke dynamics." ACM Transactions on Information and System Security (TISSEC), 2002.

[3] Bours, P. "Continuous keystroke dynamics: A different perspective towards biometric evaluation," Information Security Technical Report, 2012, pp.36-43.

[4] Bours, P., & Barghouthi, H., "Continuous authentication using keystroke dynamics," Proceedings of the Norwegian Information Security Conference (NISK), 2009.

[5] Devore, J. L., *Probability and statistics for engineering and the sciences*. (7 ed.). Thomson Higher Education, Belmont, CA, 2008.

[6] Giot, R., Dorizzi, B., Rosenberger, C., "Analysis of template update strategies for keystroke dynamics," Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011, pp.21,28.

[7] Giot, R., El-Abed, M., Rosenberger, C., "Web-Based Benchmark for Keystroke Dynamics Biometric Systems: A Statistical Analysis," Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 18-20 July 2012, pp.11,15.

[8] Gunetti, D., & Picardi, C. (2005). Keystroke analysis of free text. ACM transactions on information and systems security, 8(3), 312-347.

[9] Kang, P., Hwang, S., and Cho S., "Continual retraining of keystroke dynamics based authenticator," 2007 International Conference on Advances in Biometrics (ICB'07), Springer-Verlag, Berlin, Heidelberg, 2007, pp. 1203-1211.

[9] Killourhy, K. S., Maxion, R. A., "The Effect of Clock Resolution on Keystroke Dynamics," 11th International Symposium on Recent Advances in Intrusion Detection (RAID '08), 2008, pp. 331-350.

[11] Killourhy, K. S., & Maxion, R. A., "Comparing anomaly detectors for keystroke dynamics," 39th Annual Conference on Dependable Systems Networks, 2009, pp.125-134.

[12] Knuth, D. (1997). *Art of computer programming*, (3 ed.) Addison-Wesley, 1997, vol. 2, pp.216.

[13] Monaco J. et al., "Developing a Keystroke Biometric System for Continual Authentication of Computer Users," 2012 European Intelligence and Security Informatics Conference, 2012, pp. 210-216.

[14] Parker, D. "Windows ntfs alternate data streams," Retrieved from http://www.symantec.com/connect/articles/windows-ntfs-alternate-data-streams, 2010.

[15] Shen C. et al., "User Authentication Through Mouse Dynamics," IEEE Transactions on Information Forensics and Security, 2013, pp. 16-30.

[16] Shimson, T., Moskovitch, R., Rokach, L., & Elovici, Y., "Continuous verification using keystroke dynamics," International conference on computational intelligence and security, 2010, pp.411-415.

[17] Zack, R.S., Tappert, C.C., Sung-Hyuk C., "Performance of a long-text-input keystroke biometric authentication system using an improved k-nearest-neighbor classification method," Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), 2010, pp.27-29.

[18] Zhang, L., & Guan, Y., "Variance estimation over sliding windows," Twenty-sixth ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems, 2007.

[19] Zhong, Y., Deng, Y., & Jain, A., "Keystroke dynamics for user authentication," Retrieved from http://www.cse.msu.edu/rgroups/biometrics/Publications/SoftBiometrics/ZhongDengJain_KeystrokeDynamicsUserAuthentication_CVPR12biometricworkshop.pdf, 2012.

Evaluating Network Traffic Time and Network Latency Time for a Fault-Tolerant Multicast Routing Algorithm (FTDM)

Masoud E. Shaheen[1], Ahmed Abukmail[2] and Dia Ali[3]
[1]Fayoum University, [2]University of Houston - Clear Lake, [3]University of Southern Mississippi
Masoud.shaheen@eagles.usm.edu, Abukmail@uhcl.edu, dia.ali@usm.edu

Abstract

Due to unavoidable node and link failures, finding a fault-tolerant deadlock free multicast routing mechanism is a very important approach for building large-scale distributed memory systems. In this paper, applies the FTDM (fault tolerant multicast) routing algorithm in a wormhole-switched 2-D mesh multicomputer to evaluate two crucial performance metrics in mesh networks - network traffic time and network latency time. FTDM can tolerate convex faults without using virtual channels. This study calculates the network latency time and the network traffic time for FTDM fault tolerant multicast routing algorithm and for FT-cube2 fault tolerant multicast routing algorithm. The results demonstrate that the FTDM algorithm outperformed the FT-cube2 algorithm.

1. Introduction

In a multicomputer system, a collection of nodes (or processors) participate with each other to resolve difficult application problems. These nodes communicate their data and synchronize their efforts by sending and receiving packets through the essential communication network. One of the significant issues in parallel computing is how to send a message from one node to all destination nodes in a faulty network, where each element fails with various probabilities. The task where a source node sends a message (packet) to a destination node is called routing. Interconnection network topology is an essential element that affects multicast routing algorithms.

Many recent distributed memory systems and shared memory systems are implemented with mesh topology. These computer systems usually use the e-cube routing technique with wormhole switching because of its straightforwardness [1]. Distributed memory systems are the most beneficial architectures for building massively parallel computer systems. These systems require switching techniques to transmit messages among processors. To minimize the amount of time to deliver data, these computer systems adopt a wormhole switching technique. In a wormhole switching technique, a message (packet) is divided into a series of fixed size units of data, called flits. The information needed to choose the selection of the next channel on the route is stored on the header flit of a message and the remaining flits follow it in a pipeline fashion. Fault tolerance is a central

issue facing the design and implementation of interconnected networks for distributed-memory systems. This paper will focus on studying the performance of fault-tolerant multicast wormhole routings, FTDM, in 2D mesh networks.

2. Related work

A good fault tolerant routing should be simple, use few virtual channels, use the shortest paths between source node and destination nodes, guarantee the delivery of messages with minimum times, tolerate a large number of faults and different size of the fault region, and guarantee deadlock-free routing while minimizing disabled processors to simplify the routing algorithm. In addition, all these goals should be achieved with less consideration for hardware necessity. This paper focuses on designing a fault tolerant wormhole routing algorithm to reduce delivery time (network latency time and network traffic time). In recent years, fault tolerant routing in direct networks has been gaining attention.

In [2], Glass and Ni proposed a fault tolerant routing in meshes without virtual channels, the *negative-first*. The negative-first algorithm operates in two phases. In the first phase, the message is routed adaptively in negative direction and around faulty nodes or links, even farther west or south than the destination. In the second phase, the message is routed adaptively in positive directions to the destination. In [3], they proposed modifications to the routing logic of the base negative first algorithm to find an alternative path when blocked by a faulty component, particularly along an edge of the mesh. The algorithm is highly adaptive. However, the algorithm has a problem with large numbers of fault regions. Also, the algorithm is a unicast based multicast routing algorithm.

Youn at el [4] proposed a fault-tolerant routing method that can tolerate solid faults. The proposed algorithm requires only two virtual channels per physical channel to ensure deadlock-free meshes. The proposed scheme misroutes messages both clockwise and counter clockwise to reduce channel contention on f-rings. It is shown that the proposed algorithm is deadlock-free in meshes when it has non-overlapping multiple f-regions.

Fukushima et al [5] proposed a hardware-oriented fault tolerant routing algorithm for irregular 2D mesh networks. The proposed algorithm requires no virtual channels to ensure deadlock-free irregular 2D mesh. The proposed Position-Route algorithm for non-VC routers requires much less routing complexity. The basic idea is to integrate routing behaviors of the traditional message-based algorithm and to simplify the ring selection, but the effect of the number of fault regions is not presented. Chang and Chiu [6] proposed a fault tolerant multicast unicast-based routing algorithm, FT-cube2, in 2D meshes. In the FT-cube2, the well-known e-cube routing algorithm is improved in order to deal with multiple fault regions in 2D meshes using only two virtual channels per physical channel. In the FT-cube2 routing algorithm, normal messages are routed via e-cube hops. A message is misrouted on an f-ring or f-chain to destinations along clockwise or counterclockwise directions. This work is compared to the FT-cube2 routing algorithm.

The rest of this paper is organized as follows: Section 3 gives background of the work. Section 4 presents the proposed fault-tolerant routing method. Section 5 shows results and discussions. Finally in Section 6, some concluding remarks are made.

3. Background

This section briefly describes the fault model considered in this paper then explores multicast routing techniques.

3.1. The fault model

Many applications of interconnected networks require high reliability and availability. A large parallel computer requires an interconnected network to operate without packet loss for tens thousands of hours. Thus, these networks must employ the error control mechanism to continue operation without interruption, and possibly without packet loss, despite the failure of a component. The failure of a processing element and its associated routers is referred to as a node failure, and the failure of any communication channel is referred to as a link failure. In the fault model (figure 1), both node failures and link failures are considered. The fault model is the base for fault tolerant routing algorithms. Types of faults, structures of fault regions and processes to component failures determine the approaches to design a deadlock-free routing algorithm.

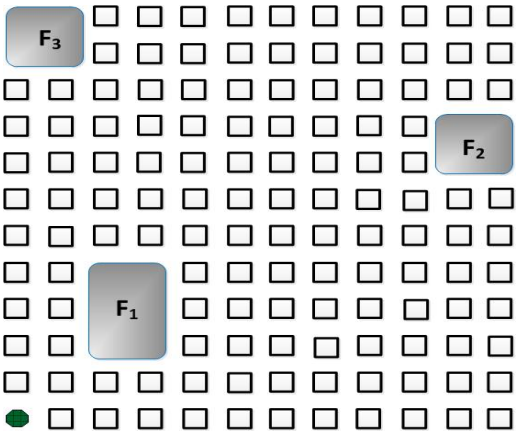


Figure 1. Fault model

In [7], the FTDM algorithm considers convex faults (also known as block faults) which are the most commonly encountered faults in mesh networks. A convex fault is a fault region such that there is a rectangle with an interior containing all and only the faulty components of the fault region and all processors, and links on its four boundaries are fault-free. A fault ring (f-ring) consists of the fault-free nodes and channels that are adjacent to one or more components of the associated fault region, F_1, as shown in Figure 1. If a fault region includes boundary nodes, the fault ring reduces to a fault chain (f-chain), F_2 and F_3, as shown in Figure 1. In the FTDM, fault information is distributed to a limited number of nodes (0, y_{bi}) in case of odd rows or (m, y_{bi}) in case of even rows) in order to avoid the fault before reaching it. Because fault information is distributed to a limited number of nodes, the FTDM is a limited-global-information-based multicasting which is a compromise of a local-information-based approach and a global-information-based approach.

3.2. Multicast routing techniques

A multicast capability delivers messages from one source node to multiple destination nodes. There are three main types of multicast routing techniques: unicast-based, tree-based and path-based [8]. The researchers construct a new type, the FTDM, which is a compromised of tree-based and unicast-based, called unicast/tree based routing. In a multicast unicast-based technique, a source node delivers a message to the destination nodes by sending multiple separate unicast messages to each destination node which needs a startup time.

The separate addressing routing is a multicast unicast-based routing, in which the source node sends directly a separate copy of the message to each destination node with startup time [9]. A multicast tree-based routing technique tries to send the message from source node to all destination nodes in a single multi-head worm

that splits at some routers and replicates the data on multiple output ports. A multicast path-based routing allows a worm to hold sorted list of several destination addresses in its header flits.

4. FTDM algorithm

In [7], a new fault tolerant multicast routing algorithm, for wormhole routed 2D mesh multicomputer was proposed. This algorithm is a multicast unicast/tree based routing algorithm. The proposed routing algorithm, called fault tolerant deadlock-free multicast, FTDM, works perfectly for convex fault regions in 2D mesh networks, f-rings and f-chains. This algorithm was proved to be deadlock free. In the FTDM no virtual channels are used. Two essential performance metrics in mesh networks, network traffic steps and network latency steps, were evaluated to measure the effect of number of destinations according to the next formulas:

Network Latency Steps (NLS) of FTDM, FTDM _Latency, is given by:

$$FTDM\_Latency=Max(LP_{(i)}, RP_{(i)}, L_4) \quad (1)$$

Network Traffic Steps (NTS) of FTDM, FTDM _Traffic, is given by:

$$FTDM\_Traffic = Traffic_{(i)} + L_4 \quad (2)$$

NLS of FT-cube2, FT_Latency, is given by:

$$FT\_Latency = Max\{ Flat_i, 1 \le i \le |D|\} \quad (3)$$

Where $Flat_i = |x_{di} - S_x| + |y_{di} - S_y| + 2*|y_{ei} - y_{bi}|$

NTS of FT-cube2, FT_Traffic, is given by:

$$FT\_Traffic = \sum_{i=1}^{|D|} Flat_i \quad (4)$$

This research measured another two important performance metrics in mesh networks, network traffic time and network latency time, as will be explained in section 5, with studying the effect of the three factors (total number of destinations, total number of fault regions and fault region size).

5. Results and Discussions

In order to test the two essential performance metrics for the FTDM, and compare it with FT-cube2 routing algorithms, a simulation study was conducted. This simulation on a 50×50 2D mesh was conducted, and double channels were used. C++ language was used to implement the two multicast routing algorithms on a PC. This section, presents the simulation results and analysis. In the simulation, wormhole routing is considered as the switching technique and the routing algorithm is also applicable with other switching techniques. The notation F represents the number of fault regions, R is the number of rows, and C is the number of columns. This configuration creates different networks with a number of processors ranging from 100 to 1080. The average number of destinations ranges from 10 to 100, and using three convex fault regions. The size of fault region ranges from 2×2 to 20×20 (number of fault nodes inside the fault region) using 25 destinations. The number of fault regions ranges from 1 to 10 using 30 destinations.

5.1. Network latency time and network traffic time analysis

In this subsection, two other important and essential performance metrics in parallel computer systems, network latency time and network traffic time, are calculated. The network latency time is the longest message transmission time involved. The network traffic time is the overall time required to deliver the message to all destination nodes [10]. From the results they are not totally independent. This means that achieving minimum network traffic time may not essentially achieve minimum network latency time at the same time, and vice versa. Network latency time depends on NLS, while network traffic time depends on NTS. The startup time also affects the value of the network latency and network traffic times. The startup time is the time acquired by the system in

preparing the message at the source node to deliver the message to the network and at the destination node to receive the message from the network. The startup time depends on the design of system software within the nodes and the interface between nodes and routers on mesh networks.

Now, network latency time and network traffic time are calculated for FTDM algorithm and FT-cube2 algorithm. The worst case of network latency time of FTDM algorithm can be calculated by:

$$FTDM\_Latency\_Time = t_{header} * D_{latency\_steps} + t_{copy} * F_{latency\_steps} + t_{channel} * FTDM\_Latency + t_{startup} * (cn+1) \quad (5)$$

Here, the time, $t_{channel,}$ is the channel time between two neighbor nodes and is multiplied by the network latency steps computed by FTDM algorithm, FTDM_Latency. The channel time, $t_{channel}$, equals the sum of the router latency time, t_r, and the channel propagation time, t_p. The time, $t_{startup,}$ is the startup time. The time, t_{header}, is the time taken to modify the message header at each destination, so it is multiplied by $D_{latency\_steps}$, which is a set of destinations participating in the longest path. Finally, the time, t_{copy}, is the time taken to copy the message at each fault region participating in the longest path, so it is multiplied by $F_{latency\_steps}$, which is a number of fault regions participating in the longest path.

The worst case of network traffic time of FTDM algorithm can be calculated by:

$$FTDM\_Traffic\_Time = t_{header} * |D| + t_{copy} * |F| + t_{startup} * (cn+1) + t_{channel} * FTDM\_Traffic \quad (6)$$

The channel time, $t_{channel,}$ is multiplied by the network traffic steps computed by FTDM algorithm, FTDM_Traffic. The time, $t_{startup,}$ is multiplied by two because FTDM algorithm requires at most two startups to deliver a message to any set of destinations, one startup time for each subnetwork of the mesh. The time, $t_{header,}$ is multiplied by all destinations, |D|. Finally, the time, $t_{copy,}$ is multiplied by |F| because FTDM algorithm requires at most |F| message replications, one at each fault region.

The worst case of network latency time of FT-cube2 algorithm can be calculated by:

$$FTcube2\_Latency\_Time = t_{startup} * |D| + t_{channel} * FT\_Latency \quad (7)$$

The channel time $t_{channel}$ is multiplied by the network latency steps computed by FT-cube2 algorithm, FT_Latency. The time, $t_{startup}$, is multiplied by |D| because FT-cube2 algorithm requires one startup time to each destination.

Finally, the worst case of network traffic time of FT-cube2 algorithm can be calculated by:

$$FTcube2\_Traffic\_Time = t_{startup} * |D| + t_{channel} * FT\_Traffic \quad (8)$$

The channel time $t_{channel}$ is multiplied by network traffic steps computed by FT-cube2 algorithm, FT_Traffic. The time, $t_{startup}$, is multiplied by |D| because FT-cube2 algorithm requires one startup time to each destination.

5.2. Network latency time and Network traffic time results

The equations from 5 to 8 were used to calculate network latency time and network traffic time for both algorithms. Figures from 2 to 7 show the results.

Figure 2 plots network traffic time for various values of average number of destinations, ranging from 10 to 100. The figure shows that, network traffic time computed by the FTDM algorithm was nearly constant as the number of destinations increased, while the traffic time computed by the FT-cube2 algorithm increased and the FTDM algorithm was less than the FT-cube2 algorithm. This is because the traffic time values depend on network traffic steps values.

Figure 3 plots the latency time for various values of average number of destinations, ranging from 10 to 100. The figure shows that

network latency time computed by the two algorithms increased as the number of destinations increased. Clearly, at a small average number of destinations, the FTDM algorithm outperforms the FT-cube2 algorithm, while at a large average number of destinations, the FT-cube2 algorithm was less than the FTDM algorithm. At a medium average number of destinations the two curves were nearly the same.

Figure 4 shows network traffic time for various values of number of fault regions, ranging from 1 to 10 and |D| is equal to 30. The figure shows that while the network traffic time computed by the FTDM increases with the number of fault regions, it continues to be significantly lower than the traffic time generated by the FT-cube2.

Figure 5 plots network latency time for various values of number of fault regions, ranging from 1 to 10 and |D| equals 30. The figure shows that network latency time computed by the FTDM algorithm decreases as number of fault regions increases, while network latency time computed by the FT-cube2 algorithm nearly constant. Also, at a small number of fault regions, network latency time computed by the FT-cube2 algorithm was less than that computed by the FTDM algorithm while at a large number of fault regions, network latency time computed by the FTDM algorithm was less than that computed by the FT-cube2 algorithm.

Figure 6 plots network traffic time for various sizes of one fault region, ranging from 2×2 to 20×20 and |D| equals 25. The figure shows that network traffic time computed by the FTDM algorithm was nearly constant with a slight decrease as size of the fault region increased, while network traffic time computed by the FT-cube2 algorithm was nearly constant and greater than the FTDM algorithm.

Figure 7 plots network latency time for various sizes of one fault region, ranging from 2×2 to 20×20 and |D| is equal to 25. Clearly, network latency time computed by the FTDM algorithm decreased as size of the fault region increased, while network latency time computed by FT-cube2 algorithm was nearly constant and less than the FTDM algorithm.

In all tested cases, network traffic time computed by the FTDM algorithm was less than that computed by the FT-cube2 algorithm.

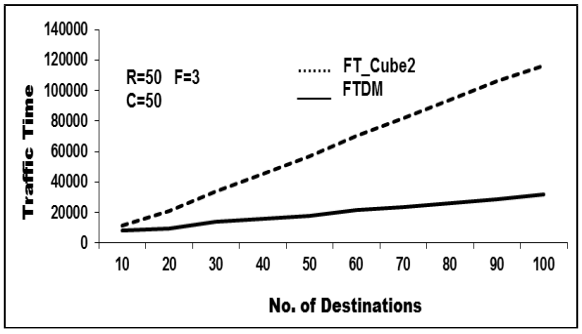


Figure 2. Network Traffic Time Vs. No. of Destinations

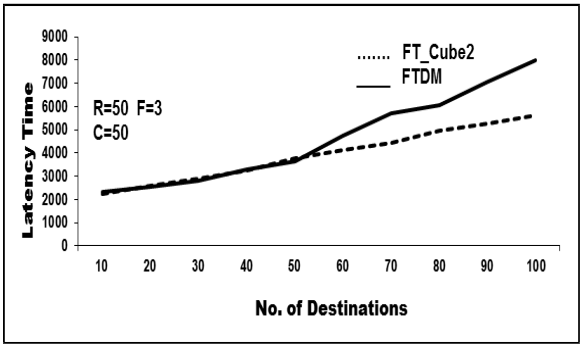


Figure 3. Network Latency Time Vs. No. of Destinations

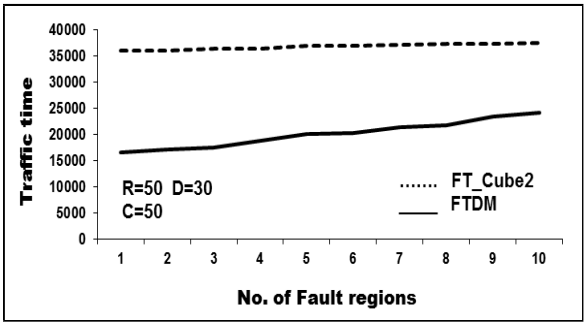


Figure 4. Network Traffic Time Vs. No. of Fault regions

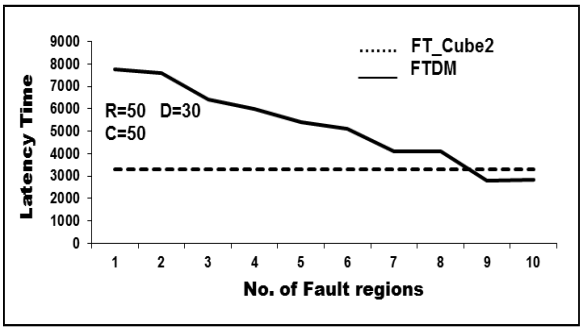


Figure 5. Network Latency Time Vs. No. of Fault regions

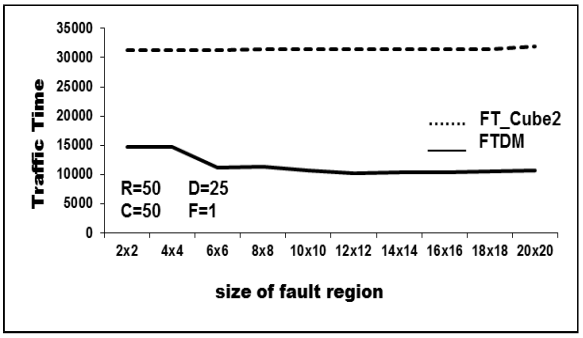


Figure 6. Network Traffic Time Vs. Size of fault regions

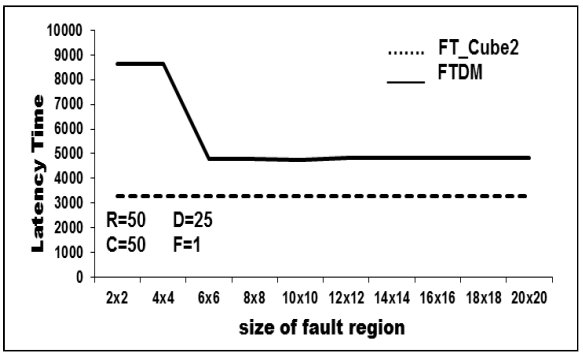


Figure 7. Network Latency Time Vs. Size of fault regions

6. Conclusions

Fault tolerant multicast routing algorithms that use less traffic and latency times are some of the most important issues that deal with the implementation of interconnection networks for large-scale parallel computers. This paper evaluated the network traffic time and the network latency time for, FTDM, a fault-tolerant multicast routing algorithm. In the case of network traffic time, the FTDM algorithm was more effective than the FT-cube2 algorithm in all tested cases.

Future work includes applying the FTDM routing algorithm on concave and irregular fault region with overlapping.

7. References

[1] R.V. Boppana and S. Chalasani, "Fault-tolerant wormhole routing algorithms for mesh networks," IEEE Trans. Computers, 44(7):848–864, July 1995.

[2] C.J. Glass and L.M. Ni, "The Turn Model for Adaptive Routing, " J. of the ACM, pp. 874-902, vol. 41, Sept. 1994.

[3] C.J. Glass and L.M. Ni, "Fault-Tolerant Wormhole Routing in Meshes Without Virtual Channels," IEEE Transaction Parallel Distributed System, vol. 7, no. 6, pp. 620-636, 1996.

[4] J.-H Youn, B. Bose, S. Park, "Fault-Tolerant Routing Algorithm in Meshes with Solid Faults," The Journal of Supercomputing, vol. 37, pp.161–177, 2006.

[5] Yusuke Fukushima, Masaru Fukushi, Ikuko Eguchi Yairi, Takeshi Hattori, "A Hardware-Oriented Fault-Tolerant Routing Algorithm for Irregular 2D-Mesh Network-on-Chip without Virtual Channels," Proc. of International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'10), 6-8 Oct., pp. 52-59, Kyoto, Japan, 2010.

[6] H.-H. Chang and G.-M. Chiu, "An Improved Fault-Tolerant Routing Algorithm in Meshes with Convex Faults," parallel computing vol. 28, pp. 133-149, 2002.

[7] Masoud E. Shaheen, Ahmed Abukmail. "A Fault Tolerant Deadlock-Free Multicast Algorithm for 2D Mesh Multicomputers," The Journal of management and Engineering Integration (JMEI), Vol. 5, No. 2, 2012.

[8] R Libeskind-Hadas, Dominic Mazzoni, and Ranjith Rajagopalan, "Tree-Based Multicasting in Wormhole-Routed Irregular Topologies," in Proc. of the Merged 12th Intl. Parallel Processing Symp. And the 9th Symp. On Parallel and Distributed Processing, Orlando, Florida, Apr. 1998.

[9] K. Hwang, Advanced Computer Architecture: Parallelism, Scalability, Programmability, McGraw-Hill, New York, 1993.

[10] R.V. Boppana, S. Chalasani, and C.S. Raghvendra, "Resource Deadlocks and Performance of Multicast Routing Algorithms, " IEEE Trans. on Parallel and Distributed Systems, vol. 9, no. 6, June. 1998.

Safety Evaluation of Buses During Rollover

Sh. M. Elseufy[1],N. M. Mawsouf[2], A. Ahmad[1]
[1]King Saud University, [2]Cairo University
sseufy@ksu.edu.sa

Abstract

This paper provides the results of a safety evaluation of buses 3D structure in the unfortunate case of rollover during an accident. Buses are a form of public transportation typically used to move people around within metropolitan cities and across cities, e.g., moving students from and to schools and moving tourists between different touristic areas.

Finite element and structure analysis techniques using LS-DYNA were applied to evaluate bus safety during rollover; and to provide design recommendations to reinforce its structure. The analysis was based on the European Economic Regulation (R66) for rollover crash test conditions and procedures. The analysis included both static and dynamic calculations and Hyperworks was used for pre and post processing. The static calculations were used to develop a bus cross section that is capable to absorb more energy from the rollover crash test and to study the complete structure and parapet bus sections. These are necessary to identify weakness areas and establish redesign recommendations to improve bus safety. The dynamic calculations were used to simulate rollover bus crashes and find residual areas with intrusion with the bus structure.

Taken together, the static and dynamic analysis indicated weakness zones (driver area, last cross member), and redesign recommendations were identified in order to provide a safe bus structure design.

1. Introduction

Bus rollover is an accident that occurs when a vehicle falls and swirls to its roof side. Rollovers accidents are common for passenger buses; Economic Commission for Europe Regulation No. 66 (ECE R66) is a standard for bus rollover testing [1, 2]. Some reasons for rollover include design weaknesses for bus superstructure, bus weight related to height, materials used, poor tests performed, etc. Bus rollovers are dangerous, where the average casualty rate is 25 per accident [2, 3].The superstructure of a bus vehicle shall be of sufficient strength to ensure no displaced part intrudes into the residual space [3]. A weak bus structure may lead to disastrous results during rollover; hence it is necessary to increase the strength of bus superstructure with minimum extra weight [4].

Previous research aimed to increase vehicle's stiffness with different techniques such as changing its material type or properties (e.g. using micro alloy instead of conventional steel [5]), or minimizing vehicle's weight by using aluminum instead of steel to reduce the overall weight [6], or using sandwich composites for the

vehicle's structures, which reduce the weight of the vehicle and also increase stiffness [7].

In addition, optimal design methods can be used to improve vehicle stiffness. These methods entail using an evaluation procedure for bus rollover crashworthiness related to vehicle weight, some examples include combining LS-DYNA and LS-OPT for finite element analysis (FEA) of bus structure [8], and designing a bus that is good at resisting rollover deformation [9]. Previous research applied tests on passenger buses to determine number and severity of casualties depending on the category of the rollover accident; which show the importance of the strength and shape of the bus's superstructure [10]. Also, experimental work was performed during rollover tests as part of the project Enhanced Coach and Bus Occupant Safety [11]. An example of a commonly applied test is the stability test, which is an evaluation that assesses steadiness of a vehicle [12].

This paper presents the results of a case study on safe bus structure design using FEA in case of rollover crash according to ECE R66. The paper proceeds by describing the test procedure, introducing regulation ECE R66, and presenting and discussing the results of the analysis. The paper concludes by providing general guidelines on how FEA can be used in lieu of actual vehicle evaluations.

2. Test procedure

The following steps were applied to evaluate bus structure design using FEA:

1) A bus structure 3D model was developed using Pro-Engineer software
2) Then, pre-processing was performed to prepare the model for analysis, i.e., creating a net of nodes and elements (mesh)
3) After that, the bus structure model was simulated under conditions from ECE R66using HyperMesh software
4) Then, the results of the simulated were analyzed using LS-DYNA software
5) Lastly, post-processing was performed to study analysis results and develop feedback reports
6) Steps 1-5were repeated until a safe bus structure design is achieved

3. Regulation R66

A rollover test on a vehicle is performed to meet all the requirements of ECE R66 [13]. These requirements include:

1. The vehicle should be stationary and all tires touching the ground uniformly
2. Loads must be kept on every passenger's seat, e.g., 75Kg is placed with its center of gravity at a height of 875mm from the floor and masses should be equally distributed
3. The vehicle is manually lifted to a uniform platform above the ground level using a crane
4. The testing apparatus should have a system to prevent tires from sliding during rollover process
5. The angular velocity of the vehicle should be zero
6. Then, the vehicle is tilted to one of its side more than 28 degrees, which leads to rollover into a ditch of depth 800mm (Fig. 1)
7. The residual space is a virtual safe area for the passengers, it should be condition tested before and after the rollover, any intrusion occurs between the residual space and the bus structure during the test will be measured by mm

These requirements are typically used for actual destructive testing on buses and other

vehicles [13, 14]. This research extends the application of these requirements using computer simulations, i.e., taking into consideration forces resulting from bus rollover in FEA and calculating bus energy during crash to describe structural behavior, strength and stiffness of a vehicle according to ECE R66. A key to computer-based evaluation of a bus rollover test is defining residual space and bus energy during crash, as they are imperative to proper application of ECE R66.

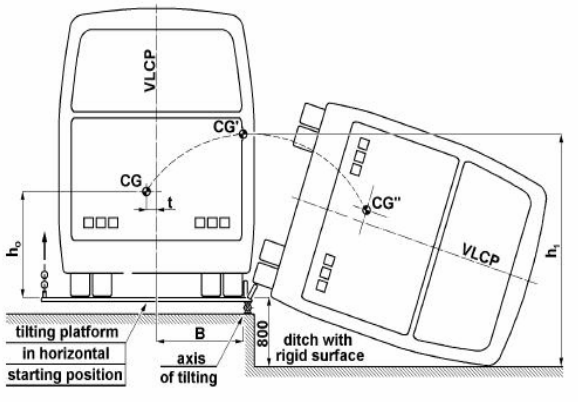


Fig.1 Rollover per ECE R66 [14]

3.1. Residual space

Residual space is the volume within bus passenger compartment which is swept when the transverse plane is moved in a straight line so that the point "R" passes from the rearmost outer seat, through every intermediate outer seat to the foremost outer passenger seats, see Fig. 2.

The position of the "S_R" point is assumed to be 500 mm above the floor under the passenger's feet, 300 mm from the inside surface of the side of the vehicle and 100 mm in front of the seat back in the centre line of the outboard seats, as illustrated in Fig.3.

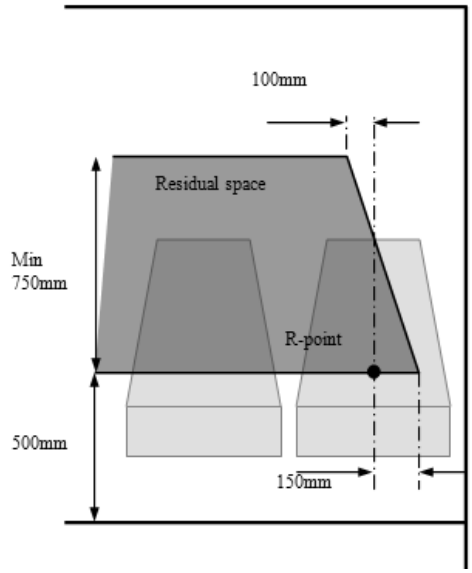


Fig.2 Residual Space per ECE R66 [14]

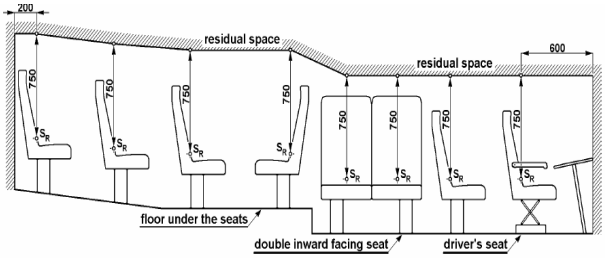


Fig.3 Longitudinal Space per ECE R66 [14]

ECE R66 aims to increase the stiffness of bus structure joints to prevent intrusion between the bus structure and its residual space. This research uses FEA to validate ECE R66 test criteria.

3.2. Bus energy calculations

The energy consumed by the body of the vehicle while impact is created during pendulum test which is mentioned in regulation by:

E=0.75.M.g.h (Nm), or

$$E = 0.75M\,g\left[\sqrt{\left(\frac{W}{2}\right)^2 + \left(H_s\right)^2} - \frac{W}{2H}\sqrt{H^2 - 0.8^2} + 0.8\frac{H_s}{H}\right] Nm$$

Where:
M= mass of vehicle (15.120ton)
g =9.8m/s^2
W= width of vehicle (m)
Hs=height of center of gravity of vehicle (m)
H= height of vehicle(1.29m)
E= 143.50 KNm (assumed per ECE R66)

This gives a maximum velocity of the center of gravity of 2.42 m/s during rollover accident [14]. This calculation will be applied using computer simulation in this research.

4. Bus rollover analysis and results

A bus model was implemented in a 3D Computer Aided Designing (CAD) model using Pro-Engineer software, then FEA used to simulate and analysis complete bus structure according to ECE R66, where about 1 million elements were used to simulate the bus structure. Different FEA elements were used to model the bus components, for exampleTRIA3/QUAD4 shell elements were used for the bus structure while bar elements and mass spots are used to simulate mechanical parts and mass distribution (e.g., engine, gearbox, axles, and welding area).

The FEA-based bus rollover analysis starts by calculating the total energy(E= 143.50 KNm) which represents the simulated forces that act on the bus structure during rollover, after that boundary conditions are added to the bus 3D model via Hypermesh software, then LS-DYNA used for solving, after that the results are analyzed using Hyperview software. Fig. 4 shows an example of the results of a deformed bus structure. The results were examined to determine bus structure weak area, and determine redesign recommendations. The FEA allows for repeating the analysis until a safe bus structure that meets ECE R66 requirements is achieved. The results of an actual bending test on specimens taken from the bus side structure were compared with a similar bending test done using the FEA to assess the validity of simulation outputs.

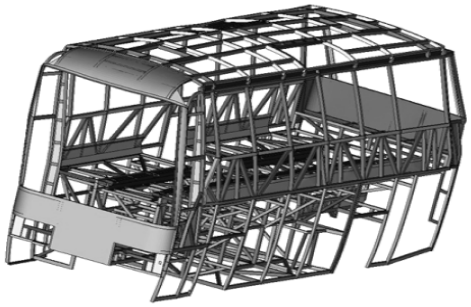


Fig.4 Deformed Bus Structure at t=200ms

The FEA analysis highlighted several weak areas in the initial bus structure, such as driver area and rear corner area. There are 3 possible ways to reinforce these areas: 1) changing the design concept for weak area, 2) modifying the thickness of the hollow tube, and 3)using different material for some parts to be stiffer. In the simulated bus rollover case, the design of some areas was modified, and the thickness for other parts was changed.

As indicated above, the results revealed that the driver area is weak, and 2 crushed tubes were obtained from the FEA, as indicated in Fig. 5. This can be dangerous to driver life, and everyone on the bus. This area was redesigned according to previous experiences and some of trial and error approach, which requires us to add steel tube40x40x2mm with length 60mm and steel plate with thickness 3mmm to reinforce the first crushed tube, and then add tube 35x35x2mm with length 350mm inside the tube 40x40x3mm and support them together by plug welding which means that thickness of both tubes become 5mm plus 3mm thickness of plate added over to be reinforce enough for the second crush tube, as indicated in Fig. 6. The FEA was applied for a second time on the redesigned bus

structure, and showed that the stiffness of the driver area was improved as indicated in Fig. 7.

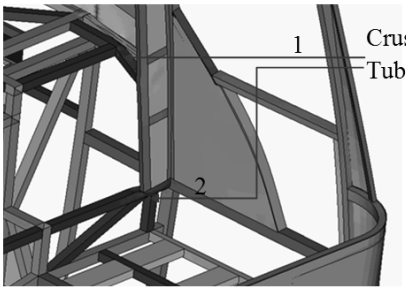


Fig. 5 Driver Weak Zone

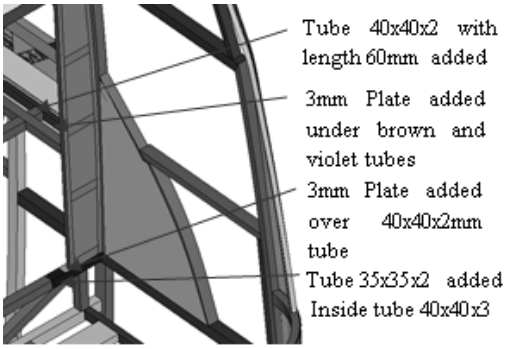


Fig.6 Driver Area Redesign

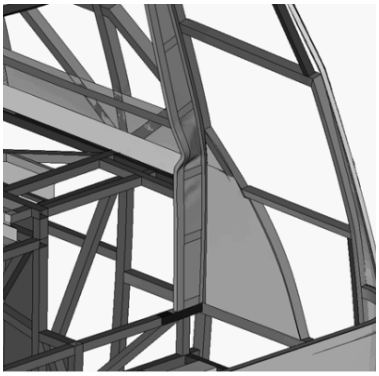


Fig.7 Results of Redesigned Driver Area

Another weak area that was revealed through the FEA is the rear corner of the bus structure, as indicated in Fig. 8, where a support tube is twisted and another is crushed, which may result in the bus structure invading the passengers' residual space, i.e., not meeting the requirements of ECE R66.The rear corner of the bus structure was redesigned as shown in Fig. 9by adding 2 reinforcement plates with 2mm thickness then add a curved tube40x40x3mmand modifying the thickness of some original support tubes to be 3 mm instead of 2 mm. The FEA was applied for a second time on the redesigned bus structure, and showed that the stiffness of the rear bus area wasimproved as indicated in Fig. 10.

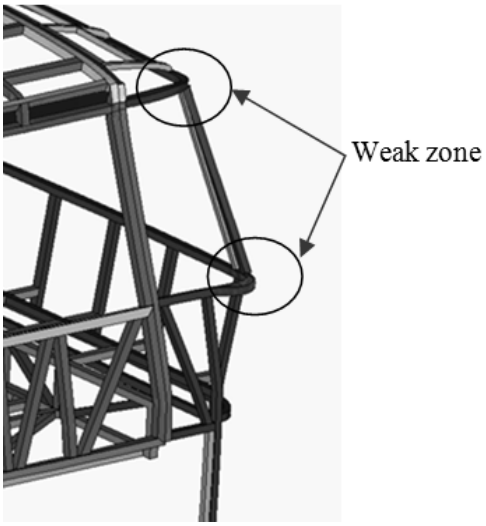


Fig.8 Rear Bus Weak Zone

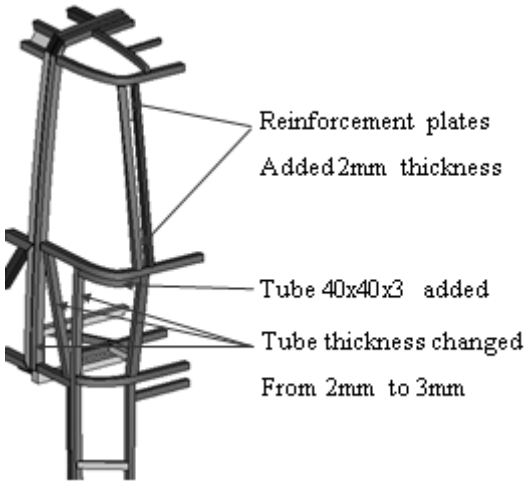


Fig. 9 Rear Area Modifications

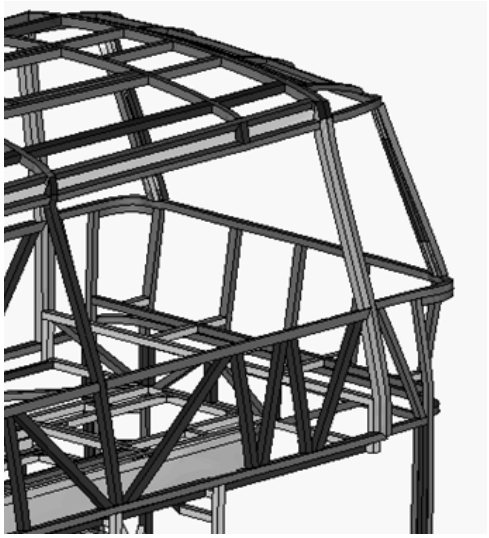


Fig.10 Results of Redesigned Rear Area

Taken together, the FEA was used as a vehicle to apply ECE R66 requirements. The evaluation was done via computer simulation of the bus energy during rollover crash. The results of the simulation revealed two weak areas in the bus structure, which endanger the life of bus driver and passengers; hence, these areas were redesigned. The FEA was applied for a second time on the improved bus design, and shows that a safe bus structure design that meets ECE R66 requirements is achieved.

5. Summary and Conclusions

A bus rollover occurs when a vehicle falls and swirls to its roof side. Reasons for rollover include design weaknesses for bus superstructure and bus weight related to height. ECE R66 is a standard for bus rollover testing. This standard is typically used for actual destructive rollover testing on buses. This paper extends the application of ECE R66 requirements via computer-based simulation.

This is achieved by applying FEA to a 3D CAD model of bus structure, and using bus energy during rollover to identify potential weak areas in its structure. The weak areas were redesigned; the FEA was applied for a second time to show that a safe bus structure is achieved.

Several lessons were learned from applying FEA to evaluate the safety of bus structures, these include:

- Saving time and money of producing a bus for each destructive test then redesign it according to test results
- Once a 3D bus structure model is produced, and can be used in computer simulation of more than one vehicle test
- FEA gives opportunity to experiment with more design choices, and the ability to test each new design

6. Acknowledgement

The authors would like to thank the Advanced Manufacturing Institute and Center of Technology Transfer at King Saud University, Riyadh, and KSA for providing the support needed to conduct this research.

7. References

[1] Commission, E. (2001). Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. Official Journal of the European Union for Legislation.

[2] Commission, E. (2010). Regulation No 66 of the Economic Commission for Europe of the United Nations (UN/ECE) Uniform provisions concerning the approval of large passenger vehicles with regard to the strength of their superstructure. Official Journal of the European Union.

[3] Matolcsy, M. (2007). The severity of bus rollover accidents. Scientific Society of Mechanical Engineers, Paper Number: 07 0989.

[4] Lan, F., J. Chen, et al. (2004). Comparative analysis for bus side structures and lightweight optimization. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 218(10): 1067-1075.

[5] Cruz, M. G. H. and A. Viecelli (2008). A methodology for replacement of conventional steel by microalloyed steel in bus tubular structures. Materials & Design 29(2): 539-545.

[6] Gauchia, A., V. Diaz, et al. (2010). Torsional stiffness and weight optimization of a real bus structure. International Journal of Automotive Technology 11(1): 41-47.

[7] Ko, H. Y., K. B. Shin, et al. (2009). A study on the crashworthiness and rollover characteristics of low-floor bus made of sandwich composites. Journal of mechanical science and technology 23(10): 2686-2693.

[8] Liang, C. C. and G. N. Le (2009). Bus rollover crashworthiness under European standard: an optimal analysis of superstructure strength using successive response surface method. International Journal of Crashworthiness 14(6): 623-639.

[9] Gao, Y. K. and F. Sun (2008). Research on the Rollover Safety of Fuel Cell Bus. Advanced Materials Research 44: 289-296.

[10] Belingardi, G., D. Gastaldin, et al. (2003). Multibody analysis of M3 bus rollover: structural behavior and passenger injury risk. Proceeding of 18th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Nagoya (Japan).

[11] Mayrhofer, E., H. Steffan, et al. (2005). Enhanced coach and bus occupant safety. Scientific Society of Mechanical Engineers, Paper Number 05-0351.

[12] Liang, C. C. and G. N. Le (2010). Analysis of bus rollover protection under legislated standards using LS-DYNA software simulation techniques. International Journal of Automotive Technology 11(4): 495-506.

[13] Automotive Industry Standards, (2008). Code of Practice for Bus Body Design and Approval. AIS-052 (Rev.1).

[14] Regulation-66, (2009). Uniform Technical Prescriptions Concerning the Approval of Large Passenger Vehicles with Regard to the Strength of their Superstructure. United Nations Publications.

Spotection: An Efficient and Versatile Parking Spot Detection System

Brandon Posey, Nitish Garg, Timothy Hall, Paulus Wahjudi
Marshall University
posey17@live.marshall.edu, garg1@live.marshall.edu, hall374@live.marshall.edu, wahjudi@marshall.edu

Abstract

Spotection is a parking spot detection and reporting system that will assist in finding parking spaces in a quick and efficient manner by providing users with a companion mobile application that provides a graphical representation of a parking lot of interest and highlights which spots are free. A lightweight image processing algorithm is utilized in combination with a video feed of the parking lot to determine which parking spots are free and to eliminate redundant elements. Along with utilizing open source software, Spotection requires minimalistic hardware and setup routines. In conjunction with the mobile application, Spotection allows for an efficient, easy-to-use, and cost-effective parking spot detection solution.

1. Problem

In today's world, automobiles are most people's main source of transportation. Within the United States alone in 2009, there were 258,957,503 registered vehicles while worldwide there are 1,554,142,411 registered vehicles total [1]. While the number of automobiles has been steadily increasing, the number of parking lots has not kept up with the increasing demand. The parking lots of many businesses, schools, government agencies and other areas are often extremely congested and become very hard to navigate during peak business hours. The "every man for himself" system that is employed today for finding parking spots is frustrating, inefficient and outdated.

The parking dilemma is very prevalent at colleges and universities alike, with so many students and faculty all searching for an empty parking spot in congested parking lots. This often results in tardiness for either the professors or the students which ultimately negatively impacts the academic experience. In addition to this, the current system of parking can also result in illegal parking in reserved spots such as handicapped or faculty parking spots.

One of the main problems with parking spot detection systems is that there is no easy way for the end user to benefit from these systems. There are plenty of solutions, but none of them center on the people who these systems are designed to help-the common driver. There may be signs in certain places showing how many free spaces are available, but what good does that do if you still have to drive around and hunt for them? With smartphone technology becoming an intricate part of our everyday lives,

why not utilize the technology we have in our hands for this purpose?

2. Current Solutions

There have been many different solutions developed for alerting people about where parking spots are located, but they are all implemented differently and they are each based on different data that is collected - mostly from third parties or from propriety sensors glued to the ground or placed under the pavement.

One such solution is from a company called Streetline; they offer a mobile application for users to install on their smartphone that pulls data in from their own proprietary sensors embedded in the pavement of each parking spot [2]. The mobile application, called Parker, can tell a user where a spot is located, set a timer for their meter, pay for parking, and also find spots according to user preferences. One of the potential problems that can occur with this system though is that all of the technology used is both propriety and expensive to install. The collective installation and maintenance costs are large enough that smaller businesses or smaller universities on a budget are unlikely to be able to implement this solution.

Another such solution is called SFpark, which is in testing in certain areas of San Francisco. This system uses sensors that are embedded in each of the parking spots just like the solution from Streetline mentioned above. SFpark also includes a mobile application for Android and iOS that is tailored to helping users find open parking spots. However, one thing that separates the solution from the rest is that SFpark can actually raise or lower the prices of parking in certain areas to encourage people to park in underutilized locations [3]. One of the issues with this system though is again the cost of the installation and infrastructure. According to the SFpark website, SFpark is currently funded by a 19.8 million dollar grant from the U.S. Department of Transportation's Urban Partnership Program. There are not many companies out there who have that kind of money. Another downfall of this approach is that the system is not very flexible. In order for the system to work, there has to be a sensor embedded in the spot. This can create a period of time when first installing the system where people will not be able to use the lot which can cause a loss of business. Also if the company were to move locations, they would have to reinstall the entire system again from the beginning. Another potential issue is the money that it costs to maintain and upgrade the sensors that are already in place. If one of the sensors dies, the monitoring will no longer work for that particular spot until a new sensor is installed.

Other solutions include mobile applications whose data is derived from a dedicated user base such as the application Open Spot from Google. The way a parking spot is identified as open is that another Open Spot user would report that the spot in question was open and it would show up on other users apps. One possible problem with an application like this is how to get the users to actively participate in parking spot reporting? The app will quickly fall flat if there are no users reporting open parking spots. Another issue would be that the application is not scalable since there has to be a certain number of people in order to make the application work.

3. Spotection System Architecture

Our parking spot detection system is called Spotection. What makes Spotection unique is that it gives the user the ability to manually identify which areas of the parking lot are parking spots. This is what allows Spotection to be lightweight as the process of determining

valid parking spots is done by the user not the system. This also allows Spotection to be dynamic and flexible to meet any user's specific needs. Allowing Spotection to be used where there are not clearly defined parking spots or even for use in helping to identify illegally parked vehicles. Figure 1 shows an outline of the entire Spotection System.

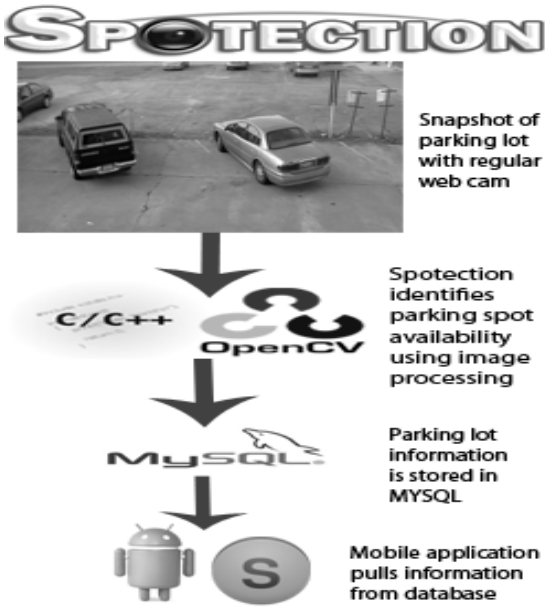


Figure 1: Spotection Monitoring System Architecture

Spotection is a cost-effective and easily-configurable parking spot detection solution. The system also includes a mobile application for use on a smart phone to help the end users make use of the Spotection system. The Spotection parking spot detection system utilizes the C++ programming language along with the Open-Source Computer Vision library (OpenCV) to process a live video feed from a webcam of the parking lot to be assessed.

While one could go into many more details as to how the system works, this paper will mainly focus on the MYSQL database and the companion mobile application sections of the Spotection System.

All the data that is gathered from the Spotection system is stored in a MYSQL database which allows the user to safely shut down the Spotection system and bring it back up without losing all the parking spots that they have already marked. This MYSQL database is the backbone of the companion Spotection mobile application.

The Spotection mobile application takes advantage of all the statistics generated by the actual Spotection monitoring system and converts them into a simple, easy to use, and convenient format for display on a mobile device.

4. Spotection Database

The real workhorse of Spotection is the MYSQL database backend of the program. This is where all the information regarding the parking lots and their respective spots are stored and where the mobile application pulls its data from. The database contains four distinct tables called parkinglots, spots, spothistory, and spotdata. Each of these tables has a unique role in the operation of the Spotection System.

The main information regarding the parking lot itself is stored in the parking lots table while all the data pertaining to the parking spots is stored in the spots table. An EER diagram of the parkinglots and spots tables is shown in Figure 2.

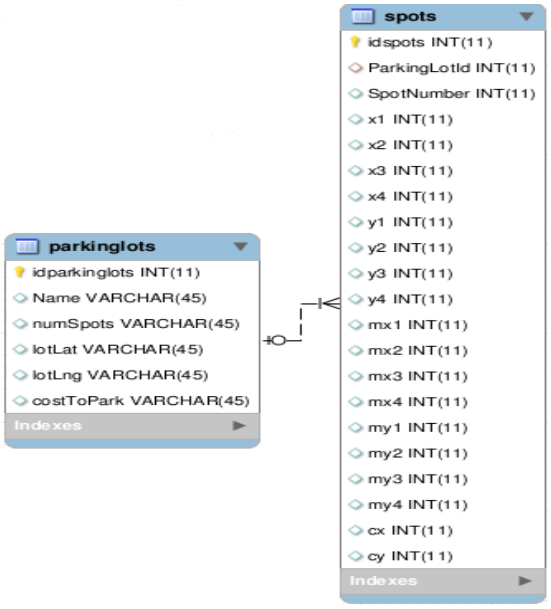


Figure 2: EER Diagram of the parkinglots and spots tables

The parkinglots and spots tables are the main tables in the database; however the spothistory table will also be a critical feature in the future. The spothistory table is the table that stores all the statistics for the parking lot so that eventually administrators can see exactly how their parking lots are being utilized and when it may be time to add more spaces to a lot. An EER diagram of the spothistory table is shown in Figure 3.

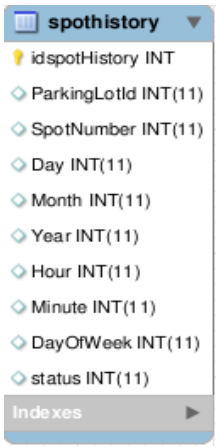


Figure 3: EER Diagram of the spothistory table

The database is the driving force behind Spotection's companion mobile application. All the information that the mobile application presents the user is gathered directly from the tables shown in Figures 2 and 3.

5. Spotection Mobile Application

The Spotection Mobile Application is in beta form right now and is written for Android which is based off of the Java programming language [5]. The mobile application is designed to be as intuitive and simple as possible while still providing the user all the information that they need in order to find a parking spot. The application allows the customer to easily select a parking lot and get all the information about that specific lot that they need. It provides directions to the lot, the cost to park in the lot (in either an hourly or flat rate format), as well as providing a summary of the parking lot by telling the user how many spots are available out of the total number of spots in the lot. Clicking on either the summary of the lot or the distance to the lot will bring up even more information like turn by turn navigation or a color coded depiction of the parking lot.

The application is built to leverage a web server to get access to the MYSQL database that is generated while using the Spotection System. However, it does not require a dedicated web server; if the user already has a website they just simply have to place the provided PHP scripts on their webserver and point those PHP scripts to the Spotection database. The mobile application will automatically search for the scripts that it needs when necessary. The actual interaction between the mobile application and the webserver however, is handled by a service that runs in the background on the mobile device and uses a combination classes from the org.apache.http Android package.

The application has three major sections designed with simplicity and functionality in mind; (1) the startup screen, (2) the parking lot selection screen, and (3) the parking lot information screen. The rest of the paper will be devoted to discussing these screens in detail.

6. Startup Screen

The startup screen is the first screen that the customer will see when they open up the application. Figure 4 is a screenshot that shows the graphical depiction of the startup screen.

The screen is a very minimalistic prompt for the customer to enter the website of the institution that is running the Spotection Monitoring System. The customer can enter either the domain name address, like www.spotection.com or an IP address like shown in Figure 4, the mobile application is built to handle both instances. The goal of the startup screen was to be very straight forward because when people are looking for a parking spot the last thing they want to have to do is open up an application on their phone and dig through a bunch of eye candy to find what they are looking for.

Figure 4: Startup Screen

While the startup screen may look simple on the outside, there is actually a lot of work being done in the background. The startup screen starts the service that will be used for all the server/application communication and then pulls in all the parking lots associated with that Spotection instance so that the lots can be displayed to the customer. After the preprocessing is done, the lots received back from the server are displayed on the parking lot selection screen.

7. Parking Lot Selection Screen

The parking lot selection screen is the screen where the customer is presented with all the parking lots that are associated with a particular Spotection System. This screen allows the customer to pick the lot that they want to search for a parking spot in. Figure 5 shows the graphical representation of the parking lot selection screen as seen by the customer. The focus of the user interface is again on simplicity and ease of use, allowing the customer to quickly find the lot they are looking for saving them precious time.

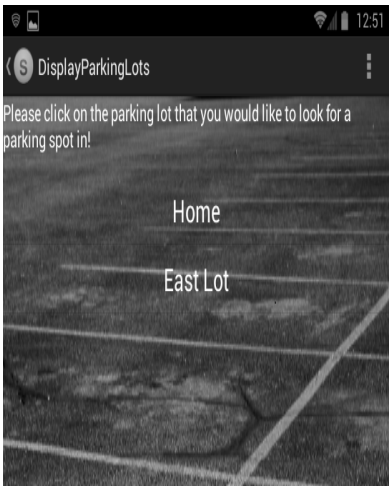


Figure 5: Parking Lot Selection Screen

When a user selects a parking lot from the list displayed on the screen, the service running

in the background first goes out and retrieves the latitude and longitude coordinates of the chosen parking lot as well as the cost to park in the lot. This information is then stored in a class called SpotectionVars.java which extends the Application interface of Android. This is done so that all other classes within the Android Application can access these variables as needed and also so that the application does not need to communicate with the server each time it needs these variables.

After the coordinates and costs are stored, the application leverages the location services integrated into the mobile device in order to retrieve the customer's location at that exact instance. This information is then sent to a script on the webserver that uses the Google Maps Public API to determine the time and distance to the parking lot from the user's current location. This information is then displayed on the parking lot information screen in an easy to read format for the customer.

8. Parking Lot Information Screen

The parking lot information screen is the main screen of the application and is designed to have all the information that the customer might need when looking for that golden parking spot. A graphical depiction of the parking lot information screen is shown in Figure 6.

Figure 6: Parking Lot Information Screen

The parking lot information screen offers a quick summary of the number of free spots in the parking lot, the cost to park in the parking lot either as a total cost or an hourly cost, and also the customer's current distance from the parking lot. This information is gathered by performing queries on the Spotection database as well as a query to the Google Map API.

The summary of the spots available is gathered by iterating through all the spots that are registered with the selected parking lot and checking the SpotOpen field in the database. If this field contains a 1 the spot is occupied consequently if the entry is a 0 the spot is free. This information will then be displayed as the number of parking spots open out of the number of total parking spots in the lot. Upon tapping the spots available, the customer will be taken to a screen that shows a color coded graphical depiction of their chosen parking lot. The green spaces indicate a free spot while the red spaces mark a spot that is already occupied. The spots will update and change color after the interval that is defined by the customer in the mobile application settings, this value ranges from real-time to every four hours. For this reason, the

spots shown are not guaranteed to still be open or occupied by the time the user gets to the parking lot. Since the Spotection system allows the company to determine the refresh interval, the information will only be as accurate as the update interval set in the Spotection system configuration. If the update setting is for every five minutes then some spots may show up as free or occupied when they really are not simply because it has not been five minutes since the last Spotection update. This could also be caused by the customer setting the refresh interval on their mobile device to be longer than real-time in order to save battery life and reduce data use.

The parking lot is displayed in row order but while the row numbers and the order of the spots are correct and to scale, the actual placement on the screen , the direction, and shape of the parking spots are not to scale. This is due partially to the diversity of Android devices and the variety of screen sizes that are available today. Using pixels for the user interface designed is not a good practice in Android due to the different screen densities; instead it is recommended that developers design their user interface using density independent pixels [4] which look the same on any screen size or density. This causes a problem with translating the coordinates stored in the spots table in the database as they are in pixel units of an 800x600 pixel display. In order to display a perfect replica of the parking lot on all devices, there would have to be a different formula for each screen size which seems wasteful. Since, in the grand scheme of things, it would not benefit the user in anyway, it was decided to draw a scaled parking lot based on screen size.

The cost to park in the parking lot is also gathered in much the same way as the number of open spots is gathered. The application uses a PHP script to query the Spotection database for the necessary information. The cost to park information is stored in the parkinglots table that was shown in Figure 1. In order to determine if the cost is a fixed cost or an hourly cost there is an F (for fixed cost) or an H (for an hourly cost) placed in front of the actual dollar amount. This information is parsed out by the mobile application and the display changes accordingly.

The distance and time from the parking lot is handled exclusively by Google Maps. The application first accesses the latitude and longitude of the device's location and then uses a Geocoder to convert the latitude and longitude into a street address that can then be sent to Google Maps for processing. After this is completed, the application then performs the same actions with the parking lot's latitude and longitude that was retrieved when the customer clicked on a parking lot from the Parking Lot Display screen. Then the number of miles and an approximation of how long it will take to get there are displayed. If the customer taps on the displayed information, then it will show the customer turn-by-turn navigation to the parking lot from their current position using the Google Maps application if it is installed. However, if the Google Maps Application is not installed then it will open an internet browser window with the directions instead.

9. Future Work

Spotection was designed to be easy-to-understand and built upon, there are a variety of attributes that can be added in order to improve performance and create a better user experience. One of these attributes is the ability for the Spotection mobile application to be able to display special types of parking spots. This includes handicapped spots, reserved spots, or places where a parking pass is needed. This would keep many users from finding a parking

spot with the system only to get there and see that it is a reserved spot in which they cannot park in.

Other attributes along those lines would be the improved detection of motorcycles, scooter type vehicles, double-parked vehicles, and improved litter and garbage detection. Smaller items of garbage and litter are already accounted for, but some larger items such as a large cardboard box can sometimes cause the spot to appear occupied.

Another feature that was partially implemented was the ability to collect statistics about the parking lot and report them to the company administrators so that they can determine how their parking resources are really being utilized. Currently, Spotection records all of the data necessary to generate these statistics by recording each time the status of a parking spot changes in the spothistory table. However, the data in this table has not yet been implemented in the mobile application or the Spotection system itself.

As for the Spotection mobile application, there are also a number of areas where it can be improved. The user interface is meant to be simple and intuitive, but it does still need a little bit of polishing before being ready for prime-time. Also the incorporation of different statistics for the selected parking lot could prove to be useful for the customer. Take for instance implementing a system that can take the history data given in the spothistory table and calculate a percentage based on previous history when certain spots will be available. This could prove valuable to a user who is in transit to the parking lot but will not arrive until later.

To continue, the Spotection mobile application is great for the end user, but what about creating an application for administrators to be able to remotely monitor their parking lots from their smart phones? This combined with branching out the application to other platforms such as iOS, Windows Phone, and Blackberry would be a large step forward for the entire Spotection Monitoring System.

10. Conclusion

In conclusion, Spotection is a lightweight, configurable, cost-effective, and user-friendly parking lot monitoring system. It is designed to be cost-effective by requiring minimal system resources and by only requiring an inexpensive camera to operate. The software is also completely open-source to alleviate the cost that is usually associated with the proprietary software used by other companies.

Spotection is ideal for small businesses and universities who want to provide their customers with a useful service but who have a relatively small budget or those who are just looking for ways to save money. The system is extremely cost-efficient and can pay for itself in the overall satisfaction of your customers. As a business or university the goals are to make the customer's experience as pleasant as possible and Spotection can help companies accomplish this. For a business to start running Spotection, all they need are a webcam and a low-end PC. Once installed, the system can then be configured to support the certain institution's needs in a truly dynamic fashion that can expand with the institution with minimal effort.

Spotection offers a truly dynamic parking spot detection solution by allowing the user to manually mark parking spots on any surface even when there are no lines defining parking spots or the lines are so faded that they are hard to pick up. This also allows Spotection to be able to assess many different shaped parking

spots such as parallelograms instead of just rectangular parking spots.

The companion mobile application allows an institution to provide their customers with the luxury of not having to waste their precious time looking for a parking spot. The mobile application provides the users all the information that they need to find a good parking spot, the number of spots free in the lot, the cost to park in the lot, as well as turn-by-turn navigation from their current location to the specified parking lot. The mobile application also provides a color coded, and graphical depiction of the parking lot allowing users to quickly glance at the screen and know exactly what row and spot number to head towards.

Together the Spotection Monitoring System along with the companion mobile application are a starter package for anyone who is looking to provide a useful service to their customers for very little money. Like all things, there is always future work to do, but even now Spotection is fully functional and ready to go as soon as installation is completed. The system is overall an easy-to-use, configurable, cost-effective parking spot detection and notification system.

11. References

[1] *Registered vehicles: Number of registered vehicles by country*. (n.d.). Retrieved from http://apps.who.int/gho/data/node.main.A995

[2]*Sensor Installation*. (n.d.). Retrieved from http://www.streetline.com/parksight/sensorsnetwork/sensor-installation/

[3]*SFpark FAQ*. (n.d.). Retrieved from http://sfpark.org/about-the-project/faq/the-basics/

[4] *Supporting Multiple Screens*. Google. Retrieved from http://developer.android.com/guide/practices/screens_support.html

[5] *Android, the world's most popular mobile platform*. Google. Retrieved from http://developer.android.com/about/index.html

Statistical Analysis of F-15 Depot's Technical Assistance Requests and Engineering Response Times

Megan A. Kane, Andrew C. Ollikainen, Johnathan C. Saxon, and R. Radharamanan
Mercer University
radharaman_r@mercer.edu

Abstract

The F-15 Engineering Group at Robins AFB has an internal goal of responding to technical assistance requests (TARs) in no more than two days. In this paper, data from January 2009 through January 2011 were captured and analyzed using statistical methods to determine if the engineering group meets their goal of no more than two days response time on all aircraft zones, and to investigate the independence or dependence of aircraft model, work-stoppage status and aircraft zone to response time. Independence tests are conducted on a data set from the F-15 system program office. These results determine that hypothesis testing of the mean response time with respect to aircraft model, structural zone, and work-stoppage status, as well as variance analysis on mean engineering response times for the same parameters is required to describe dependencies. This further testing shows that the unit is meeting its goals and identifies areas of possible optimization.

Nomenclature

α = level of confidence
c = column
E_{ij} = expected number
n = element
O_{ij} = observed frequency
P = probability
p_{ij} = probability matrix
r = row
S_{xx} = variance
t = t distribution test statistic
u_i = probability of belonging to i
μ = mean response time
v_j = probability of belonging to j
x = observation value
$\bar{x}$ = mean of observations
χ_0^2 = chi squared test statistic
Z = normal "Z" test statistic

1. Introduction

In accordance with Air Force Material Command Manual 21-1 (AFMCMAN 21-1) [1] engineering dispositions for nonconforming technical problems beyond published authority are to be submitted via Air Force Material Command Form 202 (202). These technical assistance requests (TARs) are the vehicles used by depot maintenance organizations to receive guidance and authorization from system program office (SPO) engineering to either address a deficiency in the technical data or to deviate from the accepted materials, methods or specifications contained therein. The authority to respond to TARs is crucial to maintaining Operational Safety and Effectiveness (OSS&E) [2]. Rapid engineering responses to TARs are

critical to keeping aircraft maintenance on schedule thereby maximizing asset availability.

The F-15 Engineering Group has an internal goal of responding to 202s in no more than two days. Currently metrics determine if the office is meeting the two-day goal, but no metrics are tracked to determine if certain areas and systems of the aircraft take more time to design repairs than others.

Data from January 2009 through January 2011 were captured and analyzed using accepted statistical methods to determine if the group meets their goal of a two day response time on all aircraft zones, and to investigate the independence or dependence of aircraft model, work-stoppage status and aircraft zone to response time. In this paper, it is shown that there is, at most, a minor correlation of which aircraft zone or system drove a 202 and whether the engineering response time meets the goal. The analysis also indicates that the model of aircraft is related to the number of requests logged for certain structural or system zones. This information is important in optimizing the Engineering Group to best serve the maintainers with a limited number of engineers.

2. Data parameterization

Raw data were obtained from the electronic 202 system from the F-15 system program office. The data captured include items that are incomplete, do not comply with field convention or are otherwise corrupted. Analysis of these items was not feasible, and they were rejected.

2.1. Structural/system zone

Each part numbers in the data set takes the form 68-XX with the XX corresponding to the aircraft section or system to which that part belongs. There are 45 different sub-groupings for F-15 and part numbers further grouped into 16 general systems [3, 4]. The data were grouped into four overarching zones as shown in Figure 1. Table 1 shows the data codes to zonal assignment.

2.2. Model type

The F-15C/D are air superiority fighter aircrafts that share the majority of components but differ in seating configurations with single seat configuration in the F-15C and 2 seat trainer configuration in the F-15D. The F-15C and F-15D are the older generation of the F-15 fleet having been introduced in 1979, and built until 1985 [5]. The F-15E Strike Eagle is a dual role, interdiction and air superiority fighter aircraft.

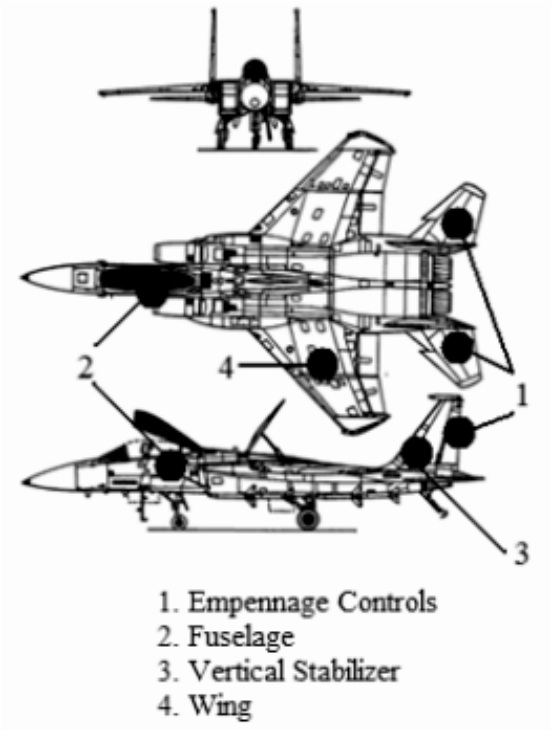


Figure 1. Illustration of F-15 zones

The F-15Es are newer aircraft, having been introduced in 1988. As seen in Table 2, the F-15E has a greater maximum takeoff weight and greater engine thrust than the F-15C/D [6].

Table 1. Data codes to zonal assignment

Code	Assignment	Zone
31	Forward Fuselage	Fuselage
32	Center Fuselage	Fuselage
33	Aft Fuselage	Fuselage
34	Reserved	Fuselage
35	Canopy/Windshield	Fuselage
36	Speed-brake	Fuselage
XX	Unknown	Fuselage
10	Reserved	Emp Cont
21	Horizontal Stabilizer	Emp Cont
20	Reserved	Emp Cont
24	Rudder	Emp Cont
23	Vertical Stabilizer	Vert Stab
11	Inner Wing	Wing
16	Reserved	Wing
17	Aileron	Wing
18	TE Flaps	Wing
19	Reserved	Wing

Table 2. F-15 model information

Model	In Service	Max. Take Off	Thrust (each)
F-15C	1979	68K lbs	23,450 lbs
F-15D	1979	68K lbs	23,450 lbs
F-15E	1988	81K lbs	29,000 lbs

2.3. Work-stoppage category

When the maintenance organization lacks engineering technical data required to do their job, and the work underway cannot be completed, they will submit a TAR with a "Work-stoppage" box checked. When the engineering group receives work-stoppage TARs, they need to be handled at high priority in a very expeditious manner [7]. Although unusual, the engineering group may move to 24 hour operations in order to support a work-stoppage 202.

3. Methodology

In order to identify what analyses to conduct on the data, several independence tests are conducted. These results of the independence testing are used to identify the appropriate hypothesis testing.

3.1. Independence testing

Tests are conducted to determine whether or not the response time is independent of the model of aircraft, the structural zone, or the work-stoppage category, and also whether or not the model of aircraft is independent of structural zone (i.e., are structural issues from one zone more prevalent on certain aircraft models than on others?). The response times are broken out into 5 groups: 0 to .5 day, .5 day to 1 day, 1 day to 1.5 days, 1.5 days to 2 days, and more than 2 days. The tests are conducted using the contingency table method [8] using a confidence level of 95% for an "α" value of 0.05. This "α" value is chosen for all subsequent testing. The contingency tables are populated with the observed frequencies. Expected frequencies are calculated using:

$$E_{ij} = n\hat{u}_i\hat{v}_j = \frac{1}{n}\sum_{m=1}^{c} O_{im}\sum_{k=1}^{r} O_{kj} \qquad (1)$$

The test statistic is calculated using:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)} \qquad (2)$$

The null hypothesis that the two classifications are independent is tested using the parameter:

$$\chi_0^2 < \chi^2_{\alpha,(r-1)(c-1)} \qquad (3)$$

If the parameter of equation (3) is true, then the hypothesis that the classifications are independent is confirmed. If any of these tests resulted in the conclusion that the data is independent of each other, there would be no cause to continue the analysis. Independence

testing indicated analyses of several relations would benefit from hypothesis testing.

3.2. Hypothesis testing

First, hypothesis tests on the mean response time for each zone were conducted, based on the organization goal of 2 days for all requests. The distribution of the response times, both overall and broken out by zone is assumed to be normal, since all samples are well in excess of 100 points. The mean and variance for each zone and for the entire set are calculated using equations (4) and (5) for the mean and the variance respectively:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (4)$$

$$s^2 = \frac{S_{xx}}{(n-1)} = \frac{[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}{(n-1)} \quad (5)$$

The standard deviation is obtained by taking the square root of the variance. Two-sided confidence intervals on the mean response time are generated using:

$$\bar{X} - t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}} \quad (6)$$

Hypothesis tests on the mean response time are conducted. The test statistic is calculated using:

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (7)$$

Next, hypothesis tests and confidence intervals calculated on the mean response time were broken out by work-stoppage category, using the same procedure as in the previous tests. Then, two-sample hypothesis tests on the mean response time were conducted, to see whether or not requests took longer for certain categories than others. Five of these tests are carried out: Stoppage vs. Non-Stoppage, Stoppage vs. Non-Stoppage within the Fuselage Zone, Wing vs. Empennage Controls, Fuselage vs. Empennage Controls, and Vertical Stabilizer vs. Empennage Controls. The first two tests are conducted as a result of the independence tests, and the latter three are conducted to clarify the results of the hypothesis testing on the structural zones. The mean and variance are calculated for each category using equations (4) and (5) respectively. Again, the standard deviation is generated by taking the square root of the variance. The hypothesis tests are conducted and the degrees of freedom for the hypothesis tests are calculated using:

$$\nu = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1 + 1} + \frac{(S_2^2/n_2)^2}{n_2 + 1}} - 2 \quad (8)$$

The test statistic is calculated using:

$$t_0^* = \frac{\bar{X}_1 - \bar{X}_2}{[\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}]^{1/2}} \quad (9)$$

Finally, the relationships between the prevalence of issues from a particular structural zone on the model of aircraft were tested. This is done using the procedure outlined in [8] specifically the large sample test. In all, 12 separate tests are conducted – for each of the 4 groups and the proportions of issues are compared between the F-15C and F-15D, the F- 15C and the F-15E, and the F-15D and the F-15E. Sample

proportions of the number of issues for each zone relative to the total number of issues for the model are calculated. Then, the estimated proportions are calculated. The test statistic for each test is calculated using equation (10). For the one-sided tests, the generic p-value comparison procedures outlined in Devore [9] are used. P-values for the test statistics are calculated using equation (11). The p-values are compared to the level of significance to determine whether or not to reject the null hypotheses.

$$Z_0 = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \quad (10)$$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{(-u^2/2du)} \quad (11)$$

4. Results and discussions

4.1. Independence testing

The results of the independence testing of response time to aircraft model and test zones, summarized in Table 3, indicate that response time is independent of the model. The differences of response time for F-15Cs, F-15Ds and F-15Es are not statistically significant. However, response time is dependent on the aircraft zone. The independence testing of the relationship of aircraft model to zone shows that the zone driving the TAR is dependent on the aircraft model. From a precursory look at the data, it seemed that the predominance of wing issues is experienced by only one or two models. As a check to better understand this relationship an independence test on aircraft model to zone was run, this time neglecting the influence of the wing. This test showed independence. From this, it could be inferred that apart from wing issues, the model of aircraft really is independent of the zone.

Table 3. Independence testing of response time to model and zone

Test	χ_0^2	df	$\chi_{\alpha, df}^2$	Ind.?
Response Time vs. Zone	45.40	12	21.03	No
Response Time vs. Model	4.96	8	15.51	Yes
Model vs. Zone	22.67	6	12.59	No
Model vs. Zone, no Wing	5.63	4	9.49	Yes

Independence testing of the relationship of work-stoppage status and engineering response time across all of the aircraft zones indicates, as shown in Table 4, that for all zones, response time and work-stoppage status are not independent. To determine which of the zones show dependencies, independence testing on the relationship of work-stoppage status and engineering response time is conducted for each of the aircraft zones.

Table 4. Independence testing of response time to stoppage category by zone

Test Zone	χ_0^2	df	$\chi_{\alpha, df}^2$	Ind.?
All Zones	12.75	4	9.49	No
Wing	7.07	4	9.49	Yes
Fuselage	14.08	4	9.49	No
Vert. Stab.	7.34	4	9.49	Yes
Emp. Cont.	5.77	4	9.49	Yes

All aircraft zones, except the fuselage, show independence to response time with respect to work-stoppage status. This served to notify the engineering group that further analysis would need to be conducted to determine the relationship of work-stoppage status and engineering response time for the fuselage zone.

4.2. Mean response time

As stated previously, the F-15 engineering group has an internal goal of answering all 202s in no more than two days. It was tested to determine

if, for any zone of the aircraft, the mean engineering response time is significantly greater than the engineering group's two days goal. The results of that test, shown in Table 5, indicate that no zones take longer than two days on average to receive an engineering response.

Table 5. Mean response time in relation to F-15 internal goal by zone (H_0: $\mu_1 = \mu_0$)

Test Zone	$\mu_1 \leq \mu_0$
All Zones	REJECT H_0
Wing	REJECT H_0
Fuselage	REJECT H_0
Vertical Stabilizer	REJECT H_0
Empennage Controls	~~REJECT~~ H_0

All but the empennage controls zone are indeed lower than the two days goal. The empennage controls group, as shown in column "AvgRT" in Table 6, has an average engineering response time of 2.19 days.

Table 6. Basic statistical information for mean response time tests by test zone

Zone	Count	Sum	SumSq
Wing	1106	1453	4759
Fuselage	749	894	3221
Vert. Stab.	366	595	2800
Emp. Cont.	204	447	2822
Total	2425	3388	13611
Zone	AvgRT	Var	StdDev
Wing	1.31	2.59	1.61
Fuse	1.19	2.88	1.70
Vert. Stab.	1.62	5.02	2.24
Emp. Cont.	2.19	9.07	3.01
Total	1.40	3.66	1.91

Although greater than the two days goal, the difference is not statistically significant to say its mean is greater. The total average engineering response time is well below the goal at only 1.4 days. Therefore, it can be asserted with confidence that the goal is being met, and in most instances exceeded.

4.3. Two-sample testing of mean response time to work-stoppage status and zones

Due to the critical nature of keeping the maintainers working on the aircraft, tests were conducted of mean response time in relation to work-stoppage status to ensure that the engineering response time for work-stoppages are more rapid than for non-work-stoppages. As seen in Table 7, non-work-stoppage TARs have greater (slower) engineering response times; work-stoppages get faster responses as they should.

Table 7. Mean response time to work-stoppage status and zone (H_0: $\mu_1 = \mu_2$)

Test Category	H_1: $\mu_1 \neq \mu_2$	H_1: $\mu_1 > \mu_2$	H_1: $\mu_1 < \mu_2$
1: Non-Stoppage 2: Stoppage	REJECT H_0	REJECT H_0	~~REJECT H_0~~
1: Non-Stoppage, Fuselage 2: Stoppage, Fuselage	REJECT H_0	REJECT H_0	~~REJECT H_0~~
1: Wing 2: Empennage Controls	REJECT H_0	~~REJECT H_0~~	REJECT H_0
1: Fuselage 2: Empennage Controls	REJECT H_0	~~REJECT H_0~~	REJECT H_0
1: Vertical Stabilizer 2: Empennage Controls	REJECT H_0	~~REJECT H_0~~	REJECT H_0

Tests on the non-work-stoppage status with respect to response time for the fuselage section were also performed. The fuselage section also followed the trend of the entire aircraft in having work-stoppages being responded to faster than

non-work-stoppages. Mean response times for the empennage controls zone showed significantly longer response times than any of the other zones of the aircraft. Further analysis would be needed to explain this, but the way the raw data is grouped may be having an impact. Also, with so many control surfaces in the empennage, and the interaction those flight controls have on other aircraft systems, it may be that engineering response take longer to develop.

4.4. Two-sample prevalence testing of issues by zone and model

In order to compare the number of TARs being submitted on each of the aircraft models, two-sample prevalence tests (p-tests) were conducted for pairs of models and sequenced them to cover each combination for comparison. As the results in Table 8 shows, the F-15D has more wing issues than the F-15C. Also, the F-15E has more wing issues than the F-15C. From the findings, the F-15D may have slightly more prevalence of issues than the F-15E, but not by a significant amount. The potential of more issues in the F-15D is somewhat expected. The F-15D is a trainer aircraft. Being a trainer, it is subjected to higher loads more frequently than the other models, especially in the wings [10].

Table 8. Two-sample prevalence-test results of wing issues by model

Model	H_0: $p_1 = p_2$	H_1: $p_1 \neq p_2$	H_0: $p_1 \geq p_2$	H_1: $p_1 < p_2$
1: F-15C	REJECT H_0		REJECT H_0	
2: F-15D				
1: F-15C	REJECT H_0		REJECT H_0	
2: F-15E				
1: F-15D	~~REJECT H_0~~		~~REJECT H_0~~	
2: F-15E				

Again, from Table 2, the F-15E has a much larger takeoff weight, and higher thrust engines. For a similar test plan and a g-rated aircraft, it would make sense that the higher wing loading, (the pressure exerted on the wings in flight) [11, 12] could lead to more instances of damage.

For the prevalence of fuselage issues being submitted in TARs, the F-15 C and the F-15D are similar. This may be from the fact that the design of the F-15C and F-15D fuselages are nearly identical rear of the aft cockpit longeron splice (Table 9).

Interestingly, the tests show that the F-15C has a higher prevalence of fuselage issues than the F-15E, but that the F-15D and the F-15E are similar. Since the F-15D and F-15C are also similar, even though statistically significant, the absolute difference of the F-15C and the F-15E must be small, as is the case. The vertical stabilizer tests showed that the F-15C had a greater prevalence for issues than the F-15D or F-15E. The test results in Table 10 show that the F-15D and F-15E are similar to each other.

Table 9. Two-sample prevalence-test results of fuselage issues by model

Model	H_0: $p_1 = p_2$	H_1: $p_1 \neq p_2$	H_0: $p_1 \geq p_2$	H_1: $p_1 < p_2$
1: F-15C	~~REJECT H_0~~		~~REJECT H_0~~	
2: F-15D				
1: F-15C	REJECT H_0		~~REJECT H_0~~	
2: F-15E				
1: F-15D	~~REJECT H_0~~		~~REJECT H_0~~	
2: F-15E				

The results of the two-sample prevalence testing conducted on the empennage controls zone for each aircraft model pairing shown in Table 11 indicate that there is no significant difference between any of the models.

The similarity of prevalence for the empennage between aircraft models is somewhat expected given that the horizontal stabilizer and rudder assemblies are the same on all of the

models [13, 14]. These assemblies are removable items that is often refurbished and replaced in depot [15, 16]. It is possible for a rudder or stabilizer to start on an F-15C, be refurbished, and then be installed on an F-15E. Having identical construction, the only differences between the stabilizer and rudder assemblies from one aircraft to others, would be caused by flight conditions.

5. Conclusions

The engineering unit is completing its TARs well below the target of two days, and work-stoppages are being addressed in an expedited manner showing that the priorities of the unit are being upheld. The response time to empennage controls takes longer than the other zones, and the F-15C has a higher prevalence of wing issues than other models—response times may be brought down by assigning another engineer to assist in empennage controls and the F-15C wing area. Although outside of the scope of this study, analysis of data down to the level of work unit code (WUC) would bring a clearer understanding of much more specific zones, and truly model-unique parts and systems.

Table 10. Two-sample prevalence-test results of vertical stabilizer issues by model

Model	H_0: $p_1 = p_2$	H_1: $p_1 \neq p_2$	H_0: $p_1 \geq p_2$	H_1: $p_1 < p_2$
1: F-15C				
2: F-15D	REJECT H_0		~~REJECT H_0~~	
1: F-15C				
2: F-15E	~~REJECT H_0~~		~~REJECT H_0~~	
1: F-15D				
2: F-15E	~~REJECT H_0~~		REJECT H_0	

Table 11. Two-sample prevalence-test results of empennage controls issues by model

Model	H_0: $p_1 = p_2$	H_1: $p_1 \neq p_2$	H_0: $p_1 \geq p_2$	H_1: $p_1 < p_2$
1: F-15C				
2: F-15D	~~REJECT H_0~~		~~REJECT H_0~~	
1: F-15C				
2: F-15E	~~REJECT H_0~~		~~REJECT H_0~~	
1: F-15D				
2: F-15E	~~REJECT H_0~~		~~REJECT H_0~~	

6. Acknowledgments

The authors thank the F-15 System Program Office and all the "Eagle Keepers" at Robins AFB for providing data and insight.

7. References

[1] United States Air Force, "Air Force Material Command Manual 21-1: Technical Order System Procedures, (AFMCMAN21-1)", 01/15/2005 AFMC, Wright-Patterson AFB, Dayton, OH, 2005.

[2] United States Air Force, Air Force Material Command Instruction 63-1201, "Implementing Operational Safety and Effectiveness (OSS&E) and Life Cycle Systems Engineering (LCSE), (AFMCI63-1201)", 10/14/2009, Wright-Patterson AFB, Dayton, OH, 2009.

[3] United States Air Force, Air Force Material Command, TO 1F-15C-4-1, "Illustrated Parts Breakdown, Airframe, USAF Series, F15C and F15D Aircraft", 1 November 2001, Change 23 - 15 February 2010, AFMC, Wright- Patterson AFB, Dayton, OH, 2010.

[4] United States Air Force, Air Material Command, "TO 1F-15E-4-1, Illustrated Parts Breakdown, Airframe," USAF Series, F15E KANE, OLLIKAINEN, SAXON 8 Aircraft, 1 July 1995, Change 34 - 1 March 2009, AFMC, Wright-Patterson AFB, Dayton, OH, 2009.

[5] United States Air Force, Air Combat Command, Public Affairs Office, "Fact Sheet: F-15 Eagle", 10/22/2009. [http://www.af.mil/information/factsheets/factsheet.asp?id=101 Accessed 4/9/11].

[6] United States Air Force, Air Combat Command, Public Affairs Office, "Fact Sheet: F-15E Strike Eagle", 10/22/2009 [http://www.af.mil/information/factsheets/factsheet.asp?fsID=102Accessed4/9/11].

[7] United States Air Force, Air Material Command, "AFMCI21-110 Robins Air Force Base Supplement 1: Depot Maintenance Technical Data

and Work Control Documents", 03/10/2006, Robins AFB, GA, 2006.

[8] Hines, W.W., D.C. Montgomery, D.M. Goldsman, and C.M. Borror, Probability and Statistics in Engineering, 4th ed., Wiley, Hoboken, NJ 2003.

[9] Devore, J.L., Probability and Statistics for Engineering and the Sciences, 8th ed., Brooks/Cole, Florence, KY 2009.

[10] Roskim, D., Airplane Design, Part VI: Preliminary Calculation of Aerodynamic, Thrust and Power Characteristics, DAR, Lawrence, KS 2004.

[11] Andreson, J.D., Aircraft Performance and Design, McGraw-Hill, Boston, MA 1999.

[12] Shevel, R.S., Fundamentals of Flight, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1989.

[13] United States Air Force, Air Force Material Command, "TO 1F-15C-3-1, Structural Repair Organizational and Intermediate, General Information", USAF Series, F15C and F15D Aircraft, 15 October 1994, Change 19 - 15 July 2010, AFMC, Wright-Patterson AFB, Dayton, OH, 2010.

[14] United States Air Force, Air Force Material Command, "TO 1F-15E-3-1, Structural Repair Organizational and Intermediate, General Information, USAF Series, F-15E Aircraft", 1 June 1996, Change 13 - 15 November 2003, AFMC, Wright-Patterson AFB, Dayton, OH, 2003.

[15] United States Air Force, Air Force Material Command, "TO 1F-15C-3-6, Structural Repair, Depot, USAF Series, F15C and F15D Aircraft", 15 November 2002, Change 18 – 15 July 2010, AFMC, Wright-Patterson AFB, Dayton, OH 2010.

[16] United States Air Force, Air Material Command, "TO 1F-15E-3-6, Structural Repair, Depot", USAF Series, F-15E Aircraft, 1 May 1996, Change 24 - 15 June 2008, AFMC, Wright-Patterson AFB, Dayton, OH 2008.

Systems Approach to Developing a Federal Government Marketing Plan

Ronald Shehane
Troy University
shehaner@troy.edu

Abstract

There is a current trend of research literature neglect on the subject of marketing to the federal government. This is especially true concerning the development of market plans that can guide businesses in achieving their sales goals with the federal government market. Market research is critical to the effort of developing marketing plans. An analysis is conducted in this study to identify the critical elements comprising a market plan and to review existing government procurement systems and the data provided that might be helpful in a market research effort. The study provides a framework of the data provided by government procurement system that will assist businesses in developing each of the critical elements required to develop a federal government market plan. The result of this study fills a gap in research literature and provides a guide for current business practice and future studies.

1. Introduction

A search of research literature indicates a significant body of academic research on market planning and strategy. However, a similar search of research literature indicates a trend of academic neglect concerning the subject of marketing to the federal government [1]. Out of a search of peer reviewed and scholarly articles, dating from 2007, none were found that dealt with planning or strategy development related to the federal government market. Coverage of the subject has been largely relegated to trade journals and magazines [1]. The federal government spends more than $500 billion per year on goods and services [2, 3]. The size of federal government market and its importance to the economy and business in general would seem to warrant more emphasis on academic research and analysis. In addition, the marketing to the federal government is very different than typical business-to-business marketing [2, 4, 5, 6]. The complexities of the government market with six hundred thousand registered competitors seeking government business, the steep learning curve and complexity involved in understanding and applying tens of thousands of pages of regulations, and the costly and voluminous proposal development required offers special marketing challenges that warrant separate study of the government market [2, 4, 6].

2. Market Plan Development

Market Research is a critical element to developing marketing strategy and an eventual marketing plan when approaching a new marketing segment, such as the federal government. As part of this marketing research a company has to evaluate the need for its

products and services in a new marketing area, assess the existing competition, and develop strategies and action plans to approach that new market segment. Such research involves assessing a company's existing business mission, current strategy, strengths, and weaknesses. However, after evaluating internal data, businesses need research information on market opportunities, competition, and customer needs to better define the new target market being approached. An effective overall marketing strategy should serve to identify what marketing research is needed and how the research should be applied toward developing a marketing plan [7].

The development of a good marketing plan is critical to the success of a business as it approaches a new market segment. The marketing plan serves to develop and direct the marketing effort. A marketing plan should address both the strategic and tactical efforts of a company in being competitive in a market segment. The strategic aspects of a marketing plan identifies the market opportunities that exist, defines the target markets within the market segment, and defines the firm's evaluation of the potential value of the market involved. The tactical aspects of the marketing plan defines the overall tactics involved in approaching the market of interest and includes how it will differentiate its product in terms of features it offers, how it will promote its product and company, how it will merchandize its product, how it will price its product to be competitive, what sales channels it will use, and what competitive services it will provide to support customers [7]. This study intends to evaluate existing federal government marketing data systems and identify how those systems satisfy the basic elements of a marketing plan as shown in Table 1.

3. Methodology

This study will consist of a content analysis of the various online procurement related government systems in existence that provide data that could serve as a basis for developing a marketing plan for doing business with the federal government. The research question addressed in this study is "What data do existing government procurement systems provide that are supportive of the market plan elements? It is hoped that the study may serve as a basis for future research and a guide to management practice in federal government marketing.

4. Government Systems Analysis

4.1 FedBizOpps (FBO)

FBO contains a reportable database of all published business opportunities over $25,000 with the federal government. It contains the detailed requirements for each procurement opportunity that are sufficient for developing and submitting a bid or proposal. The database is assessable to the government agencies, the public, and registered businesses. Businesses can register using their Dun & Bradstreet number to obtain access to all procurement documents and to have the capability to save their queries. The general public has query capability for most information including solicitations and overall requirements [8, 9].

Queries to the database can be performed for different date range by selecting a combination of federal agency or office advertising the opportunity along with a combination of product or service type code such as North American Industrial Classification System (NAICS), Product and Service Codes (PSC), or Federal Supply Code (FSC). In addition the query can be narrowed to the type of procurement such as pre-solicitation announcement, solicitation, sources sought, amendments, or award notices.

Registered businesses can specify an interest in the procurement and also specify their desire to team with other interested contractors [8, 9]. The queries and reports they produce can be saved for later reference and use.

Although the FBO system is generally viewed as a site for identifying and submitting bids or proposals for government business opportunities, it is a significant source of information needed to develop a federal government marketing plan [2, 8].

The following queries were found to be supportive of the marketing plan elements:

Market Opportunity - Search on NAIC, PSC, FSC identified existing opportunities by product or service that businesses can use to determine possible matches with their core competencies and opportunities to expand product/services to match existing markets.

Target Markets – Combination searches on NAIC, PSC, FSC along with Agency, Office, State, or ZIP selection served to identify large target areas for sales based on existing needs.

Market Potential Value – Not provided.

Product Features – The detail of each solicitation served to provide information on the features that government customers had identified. This information was provided for all search modes chosen.

Product Promotion – For all search modes the system provides the opportunity to promote a business' products/services. A business can submit proposals outlining their capability to provide products/services under a "sources sought" procurement type. A "sources sought" is a type of procurement in which the government seeks to find out the capabilities that certain types of businesses have to meet a requirement that is being considered for future solicitation. This gives businesses a chance to promote their capabilities, establish contact with procurement officials, and impact on the description of requirements eventually solicited. Businesses also have the opportunity to register as interested in the opportunity which can result in teaming offers from other vendors and further promotion of their capability within the vendor community.

Merchandizing – All search modes provide details on requirements that can be supportive of product/service design, selection choice design, product/service packaging, and display approaches that affect merchandizing. In addition, queries are provided for set-aside codes that can help special business categories to improve their merchandizing effort. The government favors certain set-aside categories such as HUBzone, Service Disabled Veteran-Owned, Woman Owned, and Small Businesses helpful to merchandizing.

Product Pricing – Not provided.

Sales Channel – The system provides information on agencies, office, state, zip, and individual contracting officers that serves to identify avenues for selling a product or service. Agency/Department searches are useful for filtering and grouping this information.

Service Support – All search modes provide details on requirements that are helpful in identifying any additional service needs that customers are requesting or that might need in supporting their main requirement.

4.2 GSA Schedules and Queries

The General Services Administration (GSA) provides two database systems that are useful for market plan development.

GSA offers an online database in its eLibrary Schedule List that provides a listing of the standard GSA schedule contracts with a description of the services/products provided to the federal government, detailed prices for each product/service, contract terms, marketing approach, and company description information. The online database provides drill down

capability so that a researcher can filter on contract category (such as Schedule 71 - Furniture or Schedule 520 – Financial Services) to see a list of vendor schedules grouped by Special Item Number (SIN) which describes the service or product categories being offered (such as Schedule 71 SIN 100 Healthcare Exam Room or Schedule 520 SIN 1 Program Financial Advisor) offered for each type, and then can select to see each vendor schedule contract with its detailed offerings and price [9,10].

GSA offers an online Sales Query Report Generation System that reports on GSA sales in a variety of ways. Report formats from which to choose are as follows: All Schedules by Fiscal Year, All Schedules by all Available Fiscal Years, SIN & Schedule Totals by Fiscal Year, All Contract Sales by Schedule by Fiscal Year, Schedule Sales Grand Total by Quarter by Fiscal Year, Total for All Quarters by Contractor by Fiscal Year, Total by Quarter & SIN by Contract Number and Fiscal Year, Total for Each Quarter for a Specific SIN by Fiscal Year, Total by Quarter & Contract for a Specific Contractor and Fiscal Year, Total by Contractor for a Specific Schedule and Fiscal Year, All Sales by Fiscal Year for a Specific SIN Number, All Contractors by Schedule by Business Size by NAICS Code [9. 11].

The following query and reporting options were found to be supportive of the marketing plan elements:

Market Opportunity - Searches on the GSA Schedule List help identify existing product opportunities by contract category and SIN for products/services. The searches assist businesses in identifying matches with their core competencies and identifying new opportunities to expand product/services to match existing product/service offerings in the government, but it fails to identify the agency or location purchasing the product/service. Descriptions of various SINs and detailed information on those services in vendor schedules help businesses refine their product/service offerings.

Target Markets – Not provided. Identifies product/services offered but not to whom they are offered.

Market Potential Value – The GSA Sales Query Reports provide sales values in the various report formats and combinations mentioned above. Market values are focused on fiscal year, SIN, NAIC, vendor, but give no clue concerning to whom and where the product/service was provided.

Product Features – The GSA Schedule List provided details of each schedule offering for each vendor for each schedule category and SIN. This provides businesses with the information needed to develop a marketing plan that matches or exceeds product features of other GSA vendors. This information can be used with the GSA Sales Query Reports by contractor to determine how well the product features offered are working for competitors in generating sales.

Product Promotion – The GSA Schedule List provides information on how competitors promote their business and products/services that can be used to match or exceed competitor efforts. This information can be used with the GSA Sales Query Reports by contractor to determine how well the promotion efforts offered is working for competitors in generating sales.

Merchandizing – The GSA Schedule List provides details on services currently offered by competitors that can be supportive of product/service design, selection choice design, product/service packaging, and display approaches that affect merchandizing. This information can be used with the GSA Sales Query Reports by contractor to determine how well the merchandizing approach used is working for competitors in generating sales.

Product Pricing – The GSA Schedule List provides detailed pricing of competitors by product/service that can be used to determine best pricing approaches. This can be used with the GSA Sales Query Reports by contractor to determine how well the pricing strategy is working for competitors in generating sales.

Sales Channel – Not provided. No information is provided on who the services is provided to.

Service Support – The GSA Schedule List provide details on additional services provided by competitors.

4.3 USASpending.Gov

USASpending.Gov is a searchable online database, accessible to the public that at no cost, that identifies federal government awards of $25,000 and above since the year 2000. Each award includes the following searchable fields: name of the entity receiving the award; agency providing the award; product/service category (NAIC and PSC); amount of the award; transaction type; funding agency; performance location to include state, city, and congressional district [3]. The database provides a graphical interface that allows users to drill down by agency, U.S. map (as shown in Figure 1), or awardee (as shown in Figure 2). The interface allows users to click on state, select award type to include contract, select product/service type, click on agency, click on vendor, click on lower level location, and finally display transaction details on each award. The system also allows the selection of a fiscal year or range of fiscal years. Reports can be displayed as trends or lists [3].

The following query and reporting options were found to be supportive of the marketing plan elements:

Market Opportunity – Searches by product/service category, agency, and location help identify existing market opportunities. The searches assist businesses in identifying matches with their core competencies and identifying new opportunities to expand product/services to match existing product/service offerings in the government.

Target Markets – Searches by location and agency within product/service category identify target markets by location and agency.

Market Potential Value – The systems provides both graphical and list views that provide sales values by product/service category, agency, and location.

Product Features – Not provided. Transaction details and product/service category provide only the broadest of descriptions of the product or service provided.

Product Promotion – Not provided.

Merchandizing – Not provided.

Product Pricing – Provides information in transaction details that shows overall dollar amount of award that can be used by a business when they lose a bid or proposal that identifies which contractor won and by how much. This is of limited use in adjusting pricing for similar proposals. This could possibly be cross matched with any available GSA contract data to compare and adjust pricing with competitors.

Sales Channel – Agency and location searches provides information on sales channels to approach within specific product type searches.

Service Support – Not provided.

4.4 Federal Procurement Data System – Next Generation (FPDS-NG)

FPDS-NG is a searchable online database, available to the general public that shows the details of contract awards of $3,000 and grater for the years 1981 to present. The system provides search and advanced search options for fields including: Agency Name, Contract ID,

Solicitation ID, Vendor DUNS number, Vendor Name, NAIC, Vendor State, Vendor City, Vendor ZIP, and Congressional District. The system provides a listing of contract awards showing amount of award, short description of requirement. When a user clicks on view, the details of the award transaction are revealed. The system provides sort capabilities on any field such as agency, date, state, ad fiscal year. The system also provides the capability of producing ad hoc reports as defined by the user [12].

The following query and reporting options were found to be supportive of the marketing plan elements:

Market Opportunity – Not useful. Searches would be of limited value since they provide only a listing of transaction details that are difficult to summarize.

Target Markets – Searches by location and agency within product/service category identify target markets by location and agency.

Market Potential Value – Not useful. Searches are of limited value since there are no summary totals.

Product Features – Not provided. Transaction details and product/service category provide only the broadest of descriptions of the product or service provided.

Product Promotion – Not provided.

Merchandizing – Not provided.

Product Pricing – Provides information in transaction details that shows overall dollar amount of award that can be used by a business when they lose a bid or proposal that identifies which contractor won and by how much. This is of limited use in adjusting pricing for similar proposals. This could possibly be cross matched with any available GSA contract data to compare and adjust pricing with competitors.

Sales Channel – Agency and location searches provide information on sales channels to approach within specific product type searches.

Service Support – Not provided.

5. Conclusions

Table 2 shows a framework of the relationships that existing federal procurement systems have with the elements required for the development of a federal market plan. The framework is based on an analysis of the query and reporting options available for each system. The framework shading indicates areas of weak or missing relationships to show shortfalls in information provided.

The framework will be a useful guide for future research and as a guide to management practice in developing marketing plans for federal government marketing efforts. Future studies of interest involve development of an integrated methodology of using systems data to develop a marketing plan, a study of the changing nature and dynamics of federal procurement systems data, an analysis of the impact government regulations and systems have had on leveling the playing field for small businesses, and identification of additional data needs for marketing efforts with the government for small businesses.

6. References

[1] Thai, K.V., "Public Procurement Re-Examined", Journal of Public Procurement, 2001, 1, 1, pp. 9-10.

[2] Bradt, J., Government Contracts Made Easy, Summit Insight, United States, 2010.

[3] USASpending.gov, Retrieved from http://www.usaspending.gov November, 2012

[4] Amtower, M., Selling to the Government, John Wiley & Sons, Hoboken, New Jersey, 2011.

[5] Santo, B., "The Fundamentals of Selling to the Feds", Catalog Age, 18, 2001, p.34

[6] Slavens, R., "Look to the Mundane for Government Opportunity", 2007 Guide to Vertical Marketing, 2007, p. 22

[7] Kotler, P. and Keller. K. , Marketing Management. 14th Ed., Prentice Hall, Upper Saddle River, New Jersey, 2012, p. 38.

[8] FedBizOps (FBO), Retrieved from https://www.fbo.gov/ November, 2012.

[9] Parvey, M., and Alston, D., The Definitive Guide to Government Contracts, Career Press, Prompton Plains, New Jersey, 2010.

[10] GSA eLibray Schedules List, Retrieved from http://www.gsaelibrary.gsa.gov/ElibMain/scheduleList.do;jsessionid=0287B64DE0E8FC8219C0E53C3B3B0FAE.node1 November, 2012.

[11] GSA Sales Query Report Generation System, Retrieved from https://ssq.gsa.gov/ReportSelection.cfm November, 2012.

[12] Federal Procurement Data System – Next Generation (FPDS-NG), Retrieved from https://www.fpds.gov/fpdsng_cms/ November, 2012.

7. Tables and Figures

Table 1. Market Plan Elements

Element	*Definition*
Market Opportunities	**Potential product needs and customers**
Target Markets	**Customers that are focus of marketing**
Market Potential Value	**Potential dollar value of markets**
Product Features	**Features of product that customers desire**
Product Promotion	**Promotional plan including sales, acceptance, competition**
Merchandising Approach	**Stimulate customer purchase with selection, packaging, display**
Product Pricing	**Competitive price determination**
Sales Channels	**Avenue through which company approaches customer**
Service Support	**Additional services offered customer to support product**

Table 2. Systems Market Plan Framework

	FBO	GSA Sched & Queries	USASpending	FPDS - NG
Market Opportunities	X	X	X	
Target Markets	X		X	X
Market Potential Value		X	X	
Product Features	X	X		
Product Promotion	X	X		
Merchandizing Approach	X	X		
Product Pricing		X	X	X
Sales Channels	X		X	
Service Support	X	X		

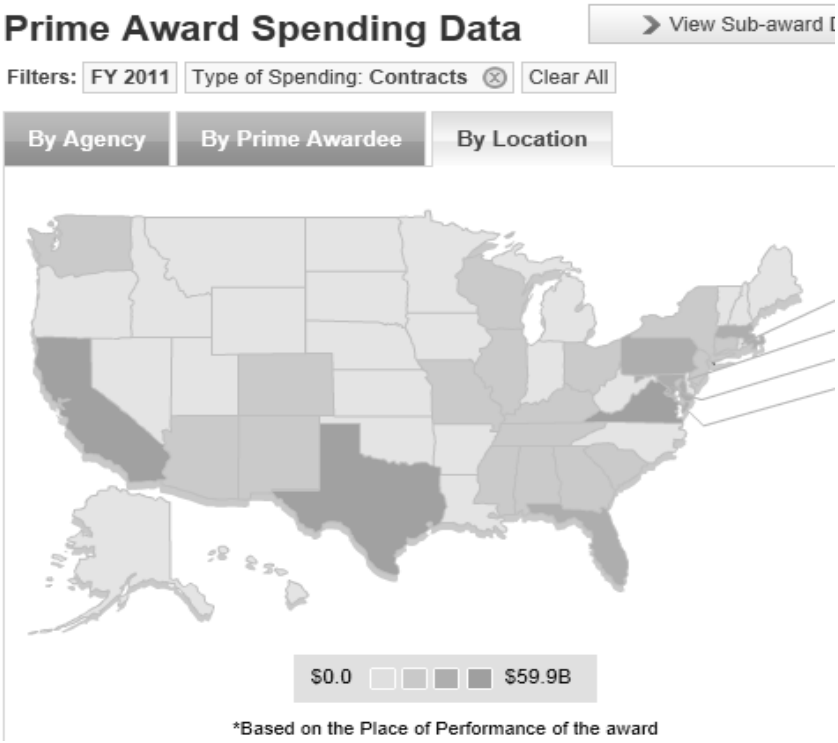


Figure 1. USASpending.Gov location drill-down

Figure 2. USASpending.Gov awardee