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

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Evaluating Pay for Performance

Shaun Trochessett, Gordon W. Arbogast
Jacksonville University
garboga@ju.edu

Abstract

Pay for Performance (P4P) programs, otherwise known as variable compensation, incentive pay, or pay at risk, are designed to incent individual or team contributors to meet certain performance goals. The best design of P4P programs has long been debated, as well as the overall impact on organizational performance, and the risk associated with loss of productivity when the interests of the company and individual contributors are improperly aligned. Proponents of P4P programs believe they are able to attract and retain key talent by offering incentives to motivate and reward stellar performance. This paper investigates the P4P model, along with the extent of its impact on organizational performance. Four P4P variables are evaluated: (1) individual-based pay; (2) sales commission programs; executive compensation programs; and (4) team-based pay programs. The paper reports that a relationship may well exist between all of these P4P compensation variables and organizational performance.

Background

Pay for Performance (P4P) involves financial rewards paid to employees whose production exceeds a predetermined standard. Frederick Winslow Taylor in the 1880's popularized scientific management and financial incentives to motivate individuals to produce at higher productivity levels. He had observed that non-incentive wage systems encouraged low productivity in the steel industry. He conducted experiments on achieving optimal performance using variable pay systems with individual incentive plans (Schachter, 1989). Terms related to P4P include "At Risk Pay", "Incentive Pay" and "variable Compensation".

Many different organizations have used iterations of the P4P model to reward employees for accomplishing established objectives. To understand the extent of its impact on operational effectiveness, the four main elements of the P4P model must first be discussed: (Ellig, 2007)

Pay Levels: P4P, (versus Fixed Pay Models), is tied to incentives for successfully making individual, team, or company performance objectives. Short-term incentives are defined as less than a year. Long-term incentives are typically measured from three to five years. However, stock options are normally extended for 5-10 years. Short-term and long-term incentive plans are usually paid in stock and/or cash.

Targeted Population: Determining individuals eligible for P4P payment is key to the design of the program. The targeted group could be an individual, team, business unit, or the entire organization. Special programs could be tailored to functions of the company ranging from the Chief Executive Officer to hourly production workers in the manufacturing facilities. C-suite positions are normally covered by short and long-term incentive programs. Conversely, the theory of P4P indicates that non-exempt employees, either working in call centers or factories, are more motivated by short-term incentives that gauge output per unit produced on an hourly or weekly basis. When the entire organization is the target for short-term or long-term incentives plan, then all management levels are evaluated by operational or financial efficiencies that fall in their area and below. "Including company performance in all sublevels of the organization

is done when it is believed appropriate that individuals understand they are part of a whole, and that it is the whole that is assessed by shareholders." (Ellig, 2008) Some employers require both the individual and his/her unit to meet operational goals in order to receive the incentive.

Targeted Performance: Performance measures are defined normally at the top of the firm. They change as you move down the organizational ranks. The Board of Directors defines the Chief Executive Officer's (CEO's) principal performance factors as achieving the strategic objectives of the entire company. In turn, the CEO uses the strategic objectives as performance requirements for his direct subordinates and they cascade down the organization. If properly executed, all employees can see how their individual performance objectives correspond to the overall objectives of the organization. "An employee can see how his or her own objectives contribute to the top organizational objectives. This assessment is referred to as *line of sight"* (Ellig, 2008). The targeted performance is assessed by relating to either past performance, projected performance, or comparing these factors against competitors or industry standards. Moreover, the appropriate design of a P4P program would have a predetermined range for effective performance. If individual or group performance falls outside of the range, variable compensation will not be issued.

Performance Measures: Companies financial statements are commonly used to determine appropriate performance indicators in the short-term, less than a year, or long-term, usually 3-5 years. The following are measures that companies frequently use when employing a P4P program: (a) **Return on assets (ROA) -** frequently used in the financial sector and capital intensive industries; (b) **return on equity (ROE);** (c) **return on net assets (RONA) - c**ommonly used by capital intensive companies at the SBU level since this ratio doesn't take into account shareholders equity; (d) **return on sales (ROS) -** Used by companies in the service or manufacturing sectors; (e) **earnings before interest and taxes (EBIT) -** Arguably one of the best measures for executives; (f) **gross profit:** Often used for measuring performance in divisions of a company; (g) **management cost ratio -** "The ratio of CEO pay to a stated financial measurement (e.g. earnings, market capitalization or sales) can be compared to ratios of comparable companies. and (h) **other key performance measurements:** These cover a myriad of business-related performance indicators (some non-financial) that include a balance scorecard, productivity and efficiency, customer satisfaction and retention, legal/environmental compliance, etc. (Ellig, 2007).

It is important to note that the lower the management cost ratio, the more expensive the cost of management. The same calculation can be made for the combined pay of the five proxy-named executives. A high ratio of CEO pay to the next highest paid in the company worries many that a team approach is missing. Ratios should not only be measured and compared with peer companies but viewed longitudinally to determine if they are improving or deteriorating." (Ellig, 2007). Over the years, some organizations have implemented P4P policies with limited success while others have reported successes. Many organizations found that individual-based P4P plans, when improperly designed, can create an internal competition that is counterproductive to the group, resulting in division, strife, and distancing from a common mission and vision. "Research done by the McKinsey Quarterly, published in 2002, shows that major companies General Electric, Nokia, Hewlett Packard, Charles Schwab, and 3M have all scrapped their pay-for-performance systems. Companies that achieve objectives of developing new businesses and managing current ones have found that it is necessary to de-emphasize pay for performance, or at least use it in a reduced fashion. The key for these companies is that its employees share the same values that the firm shares. Pay-for-performance strategies fall under scrutiny because it is extremely difficult for a company to judge who deserves a reward and how much of a reward is deserved. Companies used to rely on handing out hefty stock options to high-

performing employees. However, even this practice can be flawed." (Davidson, 2003).

Tosi et. al. conducted a meta analysis in the late 1990s that tested the hypothesized relationships between firm size, performance, and CEO pay over the 1980s and 1990s.. They reported that during that time firm size accounted for more than 40% of the variance in total CEO pay, while firm performance accounted for less than 5% of the variance (Tosi et. al., 2000). Other studies in this area have also indicated that a relationship may well not exist between firm performance and CEO pay, but this study intended to get fresh data and test for this potential relationship.

Types and Value of P4P Programs

There are five types of P4P programs: (a) individual-based pay programs; (b) sales compensation programs; (c) executive incentive compensation programs; (d) team/group based variable pay programs ; and (e) and company incentive programs (Federman, 2004).

Arguments have been made on each side concerning the value of P4P programs. Some have argued that employees exhibit superior performance when competing against each other for incentives and/or promotions. When employees focus on profits or savings due to incentives, the results often lead to higher organizational performance. In an economic downturn, variable pay can often reduce potential lay offs through reduced salary expense where a large portion of pay is placed on output measures. (Federman, 2004). Others have argued that that some employees will make less effort to achieve a common goal when they work in a group than when they work alone. This is due to phenomena known as social loafing or free riding. Employees may only focus on duties that yield incentive pay; other duties may be neglected. Also, a wide variation in the efforts of group members may lead to more intra-group conflict than cooperation. Lastly, internal wage and salary relationships can be distorted i.e. lower-skilled incentive based workers may earn more than high-skilled workers not under the plan. Research done y the McKinsey Quarterly published in 2002 showed that several major companies (General Electric, Nokia, Hewlett Packard, Charles Schwab and 3M have all scrapped their P4P systems (Davidson, 2003).

Problem Statement/Hypotheses

The basis of the study is to understand the magnitude P4P programs influence organizational performance. The result of this study is designed to either: (1) fail to reject the null hypothesis; or (2) reject the null hypothesis in favor of the stated alternative hypothesis. Given the different types of pay for performance programs, four separate hypotheses were derived.

> ***Null Hypothesis #1: (H_{O1})*** *There is no verifiable relationship between Individual-Based Pay Programs and organizational performance.*
> ***Null Hypothesis #2: (H_{O2})*** *There is no verifiable relationship between Sales Commission Programs and organizational performance.*
> ***Null Hypothesis #4: (H_{O3})*** *There is no verifiable relationship between Team-Based Pay Programs and organizational performance.*
> ***Null Hypothesis #3: (H_{O4})*** *There is no verifiable relationship between Executive Compensation Programs and organizational performance.*

Alternate Hypotheses (H1-H4): *Individual-Based Pay Programs (H_{11}), Sales Commissions Programs (H_{21}), Team-Based Pay Programs (H_{31}) or Executive Compensation Programs (H_{41}) positively impact organizational performance at the corporate and/or unit level.*

RESEARCH DESIGN & METHODOLOGY

This assessment will use research and analysis to determine the relationships of various P4P plans on performance. Individual compensation programs focus on the individual's performance and assume that the individual can control their behavior. By giving him/her a reward for good behavior, he/she may

choose to improve his/her behavior and work towards achieving a pre-established goal. If the individual meets the expectations, he/she are rewarded and the behavior should be repeated. "Among individual performance based systems we can find the following: (1) skill based systems; (2) competency based systems; (3) recognition programs; (4) spot awards; and (5) suggestion programs". (Fox Lawson and Associates, 2007).

Data

The financial data on firm performance was sourced from Standard and Poor's Compustat database, as well as Dunn and Bradstreet's Hoovers and corporate filings with the SEC as made available through EDGAR. The data set for this analysis comprised 47 of the 50 best performers in the S&P 500, as identified by Business Week magazine, based on two core financial measures: average return on capital and growth, both measured over a 36 month period (Foust, 2009). CEO pay was calculated using salary, non-equity incentive and stocks and options. It does not include indirect and deferred compensation. Options are valued using the current market price of the asset or liability (mark to market). Three companies were outliers due to CEO's taking no salary or bonus. (Apple, Google and Apollo Group) Given that current compensation is based on prior performance the 36 month performance measure more closely aligns with the compensation awards than current year performance. Statistical tests were done on 611 observations.

Individual Based-Pay Programs - *Null Hypothesis #1: (H_{01})*:

Individual compensation programs have historically been linked to organizational performance especially in the private sector. On the other hand, the public sector has been characterized by a system where pay is based on seniority and tenure. However, in recent years, the public sector has experienced a shift in its compensation practices heading towards a system that rewards individuals on performance. A recent case is the United States Postal Service (USPS). In 1995 the USPS implemented its first pay-for-performance system. This new system was implemented across its entire white-collar work force. In the 10 years after 1995, the USPS has reported a net income of $10 billion offsetting 24 years of net losses. (Jay Schuster, Weatherhead, & Patricia Zingheim, 2006) Organization performance has not only improved in the financial aspect, but other service performance indicators such as workplace safety, on-time delivery, and productivity showed major improvement in the years following its implementation.

The referenced USPS pay-for-performance system ensures that there is strong association between organizational goals, unit & departmental goals, and individual goals. "The current Postal-pay-for-performance program contains 44 unit indicators with different weight for 24 different job families covering 75,000 employees. At each organizational level, great care is given to assigning the proper weight to the indicators that employees have an ability to influence" (Jay Schuster, Weatherhead, & Patricia Zingheim, 2006). In addition, each position has a mix of corporate, departmental, and individual target goals with different weights given to each category. For example, the officer and executives have more weight on organizational and departmental performance indicators and less weight in the individual performance indicators. Each individual is measured against the pre-established goals for each category based on a 15 point rating system. Once the overall rating is given, then a link is established between the individual's score and the percentage increase in base pay; the higher the score, the higher the increase. Pay increases can range from 0 to 12%. Anything above the maximum of the range for the grade level of the employee is paid out as a cash lump-sum. "For fiscal year 2004, the overall national performance score was 9 on a 15 point-scale, and unsurprisingly, the median performance rating for 75,000 employees was also 9. The ranges were 3 to 14, and when combined with a foundation of objective performance metrics, provides concrete evidence that meaningful distinctions in performance have been made" (Jay Schuster, Weatherhead, & Patricia Zingheim, 2006) However, it should be noted that UPS took into account the following: (a)

leadership and support by top management; (b) metrics that were tied back to organizational strategy and that are concrete, measurable and achievable; (c) meaningful reward differences between the reward for high performers and non- performers; and (d) communication and involvement. "Most methods of communication were applied during the program's 10 years to ensure that everyone received and understood the message that 1) performance was tied to income, 2) the compensation system was necessary for the Postal Service's success, and 3) employees are critical stakeholders". (Jay Schuster, Weatherhead, & Patricia Zingheim, 2006).

The aforementioned study is an exemplary example of proper design and execution of individual P4P programs and its impact on organizational performance; therefore, we now reject the null hypothesis in favor of the H1 portion of the alternative hypothesis.

Sales Commission Programs - *Null Hypothesis #2: (H_{O2})*:

This section explores the hypothesis regarding the effect of Sales Commissions pay programs on firm performance. A sales commission program is one of the most common P4P programs being used to incent key personnel, teams and units within an organization. A sales commission is a sum of money paid to an employee upon completion of a task, usually selling a fixed amount (or more) of goods or services. A commission may be paid in addition to a salary or instead of a salary (Jasper, 2004). Sales commission programs are primarily used to reward the attainment of company performance objectives. Such programs are used to increase sales, lower costs, increase company profits, develop new territories and improve company profit margins. Effective sales commission programs must be in alignment with the overall business strategies, strategic objectives and sales goals of the organization (Ziegler, 2003).

A sales commission program is considered a powerful management tool that businesses can use. Sales compensation is intended to direct motivate and compensate sales personnel who significantly contribute to an organization's success. Although sales commission programs are popular there are both pros and cons associated with this type of pay for performance program that must be considered. The pros include it: (1) energizes salespeople to pursue new business aggressively; (2) rewards top performers; and (3) prevents reps from defecting to competitors. The cons include it: (1) pressures salespeople to manipulate numbers; (2) motivates them to place undue pressure on customers; and (3) may engender resentment among other company employees and divisions due to perceived special treatment (Glazer, 2008).

For medium to large companies, compensation professionals are used to create appropriate pay levels, pay mix, variable pay plan structure and the use of effective performance measures. Sales roles within a firm include; New Accounts, Blended Accounts, Existing accounts, First Line Sales Managers, Outbound Telesales, Inbound Telesales, Application Engineers, and Business Development. "For a sales job, the proportion of compensation at risk typically reflects the degree of personal selling influence a seller exerts over the customer in a buying situation. Generally speaking new Account sellers have more pay at risk than other sales roles. The pay mix is normally concentrated in three ratios- new accounts, blended accounts and existing accounts" (Colletti, 2008). The most common performance measures include: total revenue, new revenue, key sales objectives, gross profit, new accounts and renewal revenue. "To ensure that salespeople are focusing on the right sales activities, which contribute to sales results that deliver economic value to the business, top sales executives look to the use of Key Sales Objectives to communicate and reward strategic selling priorities." (Colletti, 2008). The significant relationship between sales compensation and organizational performance has been historically hard to prove for two reasons- most organizations implement new sales compensation plans at one time across an entire sales team and there are just too many confounding variables.

However, a study from an anonymous company (Company ABC) recently tested a new sales compensation plan in test territories to ascertain its effectiveness on overall performance. Company ABC's current sales compensation plan was almost entirely based on commissions earned, with limited base salary and no upside opportunity for beating company established goals. The old plan was not linked to pay for performance philosophies. The new sales compensation plan changed from pure commission tied directly to total revenue to one tied to quota achievement. It offered a small base salary and the opportunity for variable pay once 80% of the sales goal was achieved. Payment acceleration was offered for each incremental percentage point of revenue achievement over goal resulting in two to three times the normal compensation rate. Stakeholders in the plan were looking for at least a one percent incremental increase in revenue. After six (6) months the test program results showed: (a) revenues were 3% higher in the territories than revenues from the old plan; (b) the net increase in revenue was 2.7 million; (c) net margins remained constant – extra revenue was real and meaningful profit. Company ABC realized the 3% sales increase was a real breakthrough. Also the new plan created a more dynamic and passionate sales culture within the organization. Overall organizational performance increased and as a result, Company ABC implemented the new sales plan companywide (Colletti, 2008). Designed and implemented correctly, commission programs appear to incent sales professionals, consequently increasing firm sales revenues. Based on this study there is evidence to support the rejection of $H_{O2.}$

Team/Group-Based Variable Pay Programs, - ***Null Hypothesis #3: (H_{O3})*:**

Concerning this hypothesis U.S. firms normally depend on individuals for meeting corporate goals. The top-down corporate structure has supported and rewarded individuals for many years. In the mid-1990s some firms started to experiment with team-oriented environments. Such a transition represented a complete cultural change, Americans having been rewarded for their individual successes from a very young age. However, for a whole group to be successful, each individual must work together, as a team to achieve success. Many organizations in the business world are changing their mentality and are utilizing teams to attempt to improve customer satisfaction, quality, and productivity (Palmieri, 2009).

There are several different types of teams within organizations. The four most common types and the most common ways of rewarding them are: (a) **Parallel Teams -** Gainsharing involves sharing with the organization the financial benefits of performance improvements (Flannery, 1995); (b) **Project teams -** A diverse group of workers combine their efforts to develop innovations and fulfill customer requirements for special projects; (c) **Work Teams -** Self-contained work units control the process in manufacturing organizations to produce a more efficient product or service; and (d) **Management Teams -** Non-executive teams that are used to provide direction for an entire organization and are responsible for the overall performance of the company. Many companies have used different types of teams within their organizations. "They are doing this in order to share, capture, and deploy the intellect and knowledge needed to gain a competitive advantage within their industries" (Kennedy, 2003). When done well, the team results include higher quality and productivity, increased innovation, higher flexibility and adaptability, and greater speed and better response to the needs of customers.

A recent study of hospitals was reported in the New England Journal of Medicine, stating "Pay-for- Performance based teams in hospitals showed a greater improvement in all composite measures of quality for patient care versus hospitals which did not institute the pay-for-performance system" (Lindenauer, 2007). Although there have been some failures of instituting team based-pay-for performance, the results seem to point in the direction that they can be very successful. Organizations that have instituted the team based system and have been able to mold the corporate culture for this type of organization have reportedly been successful. The Lindenauer study does provide quantitative

data to support the rejection of H_{O3}. In addition, there is overwhelming qualitative information that team based pay programs may well play a significant relationship in organizational performance (Kennedy, 2003). Based on this rationale it can be surmised that H_{O3} may very well also be rejected.

Executive Compensation Programs, - *Null Hypothesis #4: (H_{O4})*:

This section explores the null hypothesis regarding the effects of executive compensation on firm performance. Many Boards of Directors in setting executive compensation are faced with the principal agency problem first described in a study of CEO compensation by Kevin J. Murphy in 1986 [Murphy 1986] In this study Murphy held that the executive (as the agent), seeks higher compensation in some cases when it is not necessarily in the best interest of the principal, or in this case the shareholders. The Board of Directors is challenged to find a way to compensate the CEO in such a way as to incent behavior in the executive that will increase shareholder wealth, thus aligning the interests of both the principal and the agent. Murphy concluded in his study that "the level of managerial effort will depend on an executive's incentive contact," (Murphy, 1986). It is this theory that has fueled the exponential growth of contingent compensation, in particular stock based compensation. In 1980, 30 percent of CEOs received options, while in 1994 this figure had increased to 70 percent [Hall and Liebman, 1997]. This part of the study was an attempt to see if more recent data was consistent with previous work done in the 1990s such as the Tosi research referenced earlier (Tosi, et.al., 2000).

Given recent market performance, and the current outrage over high CEO compensation, current executive compensation levels was evaluated to see if they have a positive effect on firm performance. CEO pay was measured on salary, bonuses paid that year, as well as any options exercised that year. It did not include indirect and deferred compensation. Executive compensation was analyzed along with the market value of the firm in order to determine the impact of CEO compensation on shareholder wealth (surrogate for firm performance). CEO performance was also measured against EBIT to determine the impact on firm performance.

RESEARCH RESULTS AND CONCLUSIONS

H_O was previously rejected in favor of H1, H2, and H3. The remaining hypothesis to be tested was $H_{O4.}$ Descriptive statistical methods, qualitative factors and preexisting research were used to suggest whether or not there was anything substantive that could be derived concerning executive pay. If H_{O4} can be rejected, this would provide some evidence that variable compensation or P4P, if managed properly, may well have a positive impact on organizational performance. The alternate hypothesis is whether executive compensation has a positive impact on corporate performance. This hypothesis was evaluated using two performance measures: market value and EBIT. The market value used was the share price at closing multiplied by the number of shares outstanding at that point in time. EBIT was measured at the most recent twelve (12) months ended January 31, 2009. All tests were conducted using an alpha of .05.

Hypothesis #4A: No link exists between firm performance, "market value," and CEO compensation.

In testing H_{O4A}, a correlation test was first conducted to determine if a relationship existed between CEO compensation and market value of the firm. This resulted in a correlation value of 0.25956, indicating a small positive relationship. Since the direction of the relationship was known, one-tail tests were used. An F test was next used to determine if the mean variance between CEO compensation and market value are equal. This test resulted in a P (F<=f) one-tail of 1.907E-162. Since this is significantly less than .05, the null hypothesis was rejected and it can be assumed that the mean variances are unequal. Given this information a t test (Two Sample Assuming Uneven Variances) was next conducted. Using the standard t formula this test resulted in P (T<=t) one-tail of 6.56E-05 and t stat of 4.17486

relative to t critical of 1.67866. These results support the rejection of H_{O4A}, and conclude that there may well be a positive link between firm performance expressed as "market value" and CEO compensation.

Table 1: t-Test: Two-Sample Assuming Unequal Variances
Mkt Value vs. CEO Compensation

	Variable 1	*Variable 2*
Mean	20248.93617	7.916851064
Variance	1104787336	29.07593074
Observations	47	47
Hypothesized Mean Difference	0	
Df	46	
t Stat	4.17486041	
P(T<=t) one-tail	6.56255E-05	
t Critical one-tail	1.678660414	
P(T<=t) two-tail	0.000131251	
t Critical two-tail	2.012895567	

Hypothesis #4B: There is a verifiable link between firm performance, EBIT and CEO compensation.

In testing H_{O4B} a correlation test was first conducted to determine if a relationship existed between CEO compensation and EBIT of the firm. This resulted in a correlation value of 0.242278, indicating a small positive relationship. Next, one-tail tests were again utilized. First an F test was conducted to determine if the mean variance between CEO compensation and EBIT value were equal. This test resulted in a P (F<=f) one-tail of 1.8E-116. Since this was significantly less than .05, the null hypothesis was rejected and it was assumed that the mean variances are unequal. A t test (Two Sample Assuming Uneven Variances) was next conducted. This test resulted in P (T<=t) one-tail of 9.7E-05 and t stat of 4.05214 relative to t critical of 1.67866. Both of these results (H_{O4A} and H_{O4B}) lead to the conclusion that there may well be a positive link between firm performance expressed as EBIT and CEO compensation. While this result indicates that there may be a relationship, it is recognized that it would go counter to the robust meta-analysis done on earlier data by Tosi. More testing may need to be done to be able to gain a degree of confidence that this in fact the case.

Table 2: Summary of Research Results	Pass Alternate Hypotheses?		
	Yes	No	Undetermined
Executive Compensation Plans			√
Sales Commission Plans	√		
Team-Based Compensation Plans	√		
Individual-Based Incentive Plans	√		

RECOMMENDATIONS

Concerning Executive Compensation Plans, more testing should be conducted to determine if there is a link with firm performance and/or market value. In lieu of any additional information and particularly in view of the robust Tosi study, caution should be exercised in surmising that there is a link between Executive Compensation and firm performance and/or market value.

Any P4P programs that are implemented must be specific, measurable, attainable, realistic and timely to ensure both the employees trust the program and performance goals are attainable. All P4P plans must have a meaningful difference between the reward for high performers and non- performers to make it worthwhile for employees to try to achieve or exceed the established goals. In implementing a proper P4P plan there should be a cascading effect between organizational goals, departmental goals, and individual goals in order to support the overall strategic and financial objectives of the firm. All employees should see how their individual performance objectives correspond to the firm's strategic objectives.

REFERENCES

Beyerlein, M., Beyerlein S. & Johnson, D., Advances in Interdisciplinary Studies of Work
Teams, Volume 9. JAI Press, New York, NY. (2003).

Colletti, Jerome A. and Davis, Jay M., "Harnessing the Power of Sales Compensation: An
Analysis of 2008 Suvery Practices". Workspan, World at Work, Scottsdale, AZ.
(August 2008).

Davidson, B. "Pay For Performance Strategies: Vastly Overused". (Retrieved 03 25,
2009), from http://web.syr.edu/~rodavids/PayForPerformance.htm.

Ellig, Bruce R. The Complete Guide to Executive Compensation. McGraw Hill, NY,NY. (2007).

Ellig, B. "What Pay for Performance Should Measure". World At Work. Employee Rights. (n.d.)

Retrieved 4 2 2009 from http://www.mypersonnelfile.com/wp_commissions.php, (2008).

Federman, D. Pay-for-Performance: From theory to Reality. Workspan, World at Work, Volume 47, Number 4. Scottsdale, AZ. (2004)

Flannery, T. People, Performance, and Pay: Dynamic Compensation for Changing Organizations. The Free Press. New York, New York, (1995).

Foust, D. The BusinessWeek 50. BusinessWeek. (April 6, 2009)

Fox, Lawson, & Associates. "Pay for Performance Systems: Individual Programs". World at Work. Scottsdale, AZ. (2007).

Glazer, David, Sales Commission versus Product Performance in Financial Services. Vdm Verlag Publishers, India. (2008).

Hall, Brian J.; Liebman, Jeffrey B. "Are CEOs Really Paid Like Bureaucrats?" National Bureau Of Economic Research Working Paper Series, No. 6213. Cambridge, Mass. Oct. 1997.

Jasper, Margaret C. Employee Rights in the Workplace. Oxford University Press, Oxford, UK. (2004).

Kennedy, F. "Managing a Team-Based Organization- A Proposed Strategic Model". Advances in Interdisciplinary Studies: Team Based Organizing. Elsevier Ltd, London UK. (2003).

King, J. "High Performance Work Systems and Firm Performance". Monthly Labor Review. Vol. 118 No.5, (1995).

Lindenauer, P. "Public Reporting and Pay for Performance in Hospital Quality Improvement", New England Journal of Medicine, (2007).

Murphy, K. J. "Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts," The RAND Journal of Economics, 17, 1, (Spring 1986).

Palmieri, Dewey. "Team-Based Incentive Compensation is Vital to Successful Operating Turnarounds". Journal of Private Equity, Vol. 12, No. 3. (Summer 2009).

Platt, R. K. "Merit-Based Pay". World At Work . Scottsdale, AZ. (2003).

Platt, R. K. "Team-Based Pay". World at Work , Scottsdale, AZ. (2004).

Rhodes, Marcia. "Bonus Program Practices". World At Work . Scottsdale, AZ. (January 2009)

Schachter, Hindy Lauer. Frederick Taylor and the Public Administration Community- A Re-evaluation. State University of New York Press, Albany. NY. (1989)

Schuster, Jay P., Weatherhead, P., & Patricia Zingheim, P. "Pay For Performance Works: The United States Postal Service Presents a Powerful Business Case." *World At Work* . 1st Quarter 2006 Volume 15 Number 1. Scottsdale, AZ . (2006).

Tosi, H., Werner, S., Katz, J. and Gomez-Mejia, L., "How much does performance matter? A meta-analysis of CEO Pay Studies", Journal of Management, Volume 26, No. 2, Thousand Oaks, CA., (2000).

Yanadori, Y., & Marler, J. "Strategic compensation: Does business strategy influence Compensation in High-Technology Firms?" CAHRS Working Paper Series (2003).

Ziglar, Zig. Selling 101. Zig Ziglar Publishing Corporation. Nashville, Tenn. (2003)

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Educational Learning Software: Experiences Creating an Intuitive Software Interface

Jorge E. Arias
Florida Institute of Technology
jarias@my.fit.edu

Daniela M. Natali
Florida Institute of Technology
dnatali@my.fit.edu

Abstract

This research seeks to improve the usability features in a student's progress tracking application. The progress tracking application was developed to improve students and educator's progress measures and to help them manage their grades efficiently and effectively. As result of the study, experimenters found that the software has a strong foundation for an intuitive application and that improvements need to be made in the wording of the sections in order to match users' mental model. This paper explains the usability study conducted in order to implement a more intuitive web site.

1. Introduction

Educational institutions are struggling to increase student progress. Many new ideas have been developed in order to help educators measure the progress of students based on learning programs, and teacher effectiveness and efficiency. Progress tracking applications with high utility value have been introduced to the market place, but most of these applications lack of usability factors such as learnability and memorability. Other developers have achieved the learnability and memorability factors by cutting on the application's utility factor. The team in charge of this experiment has managed to identify user needs and to document users' requirements. Also, they have developed alternative designs that meet those requirements and have picked the best design. Currently, the experimenters are in the process of evaluating the data entry portion of the application and the user experience the application offers.

A usability study was conducted to identify usability problems as well as recommendations for improvement for the data entry section of the application. The usability study was performed on participants, between the ages of 18 and 55, recruited from a University consisting of college students, staff, and faculty. During the experiment, the experimenters pre-surveyed the participants to learn their level of expertise with educational learning software, gave the participants general guidelines, asked the participants to perform the tasks, and asked the participants to fill the post-test survey in order to gather additional information. The study performed is the first usability study that will help the software development team produce a more intuitive user interface by creating an interface that users can predict and that matches the users' mental model.

2. Literature Review

When a salesperson says a software is easy to use, it probably is for somebody. Developers usually have a group of people in mind when they come up with a product [13]. The ability of a product to match the mental model of a bigger audience could easily become a factor to the success of the product. That ability to match

users' mental models is called usability. "Usability has been defined as a measure of the quality of a user's experience when interacting with a product or system – whether a website, a software application, mobile technology, or any user-operated device" [2]. In other words, the users are the ones who decide whether a product is easy to use, satisfying, visually appealing, or rewarding. For this reason, it is essential to conduct usability testing with the potential users of a product. Overall, one can compare an interactive product, like the progress tracking application, to a play, where extensive rehearsals are expected especially close to the opening night [3].

"With more and more systems, the user is becoming the focal point of the use of the computer" [5]. In the past, users only had one system to turn to for each of their needs. Today, users have many systems to pick from in order to accomplish their tasks. Because of the surplus of products that accomplish the same tasks, users' product of choice is constantly defined by the usability factor of the product. As a result, human engineering has become a core aspect of all successful products.

The domain of the progress tracking product is student learning. In order to create a study that is both efficient and effective in assessing our product usability needs, one may consider the key usability factors of service-oriented web sites for students recommended by Robert Nathan's study:

- Emphasis on colors and fonts
- Be clear in defining the goals and main objectives of the web site
- Download speed is very important
- Navigational aspects of the web site are important but should not be overemphasized
- Interactive tools that would involve students in the web site (e.g. real-time chat, online forums, shout-boxes and blogging) are highly encouraged [6]

In addition to measuring the usability of the above key usability factors, one can expand the study to cover task completion time and error count logging. Also, it is important to do as much data collection as possible by the experimenters [8].

On top of the previous factors, there is another aspect that a salesperson can definitively use in their pitch. That aspect is a factor difficult to measure. Only the user can grade this factor. It is the factor of pleasure of the website. Website pleasure has been defined as the measure of the enjoyment of a user's experience when visiting a website. Furthermore, website pleasure is considered a prerequisite for a successful site [9].

The progress tracking software is an application in a domain where many companies are battling for market share. Those who are capable of excelling both, the previously discussed, usability factor and utility value will earn this market share.

The utility value of student learning software has a lot to do with the ability to do content management. The objective of content management is to control the potentially complex workflows through the information lifecycle [11]. As a result, the usability study performed shall include tasks that evaluate the ease of use of the information management of the website. Below are some of the heuristics recommended for testing content management usability:

- Users understanding of how the various screens work
- Which features and functions users find difficult to use
- How well the screen behavior matches users' expectations
- Which features and functions users like and do not like [10]

Another factor that this study will address is whether the language used in the content of the application by the developers is appropriate for

the audience this study is targeting. Participants who are not professors will be considered the least skilled population of users the application will encounter in the real world. If a sentence or word seem unfamiliar to the least skilled participants, the application should not be using it. A website that follows the mental model of the user should adapt the screen dialog to the users' vocabulary. Thus, minimizing the use of technical jargon [4].

This literature review summarizes the focus of the usability study. In the following sections, the performed study will be described.

3. Methodology

The usability study was conducted because the progress tracking application requires being a very easy-to-use, intuitive application. There are different ways of collecting data and conducting website usability testing. This research implements the traditional usability testing procedures: the user testing the program, the collection of data through pre-surveys and post-surveys, and collection of data through the experimenter's observations. In order to prepare the study, the team identified the tasks that needed to be examined. To control user variety, the team randomly selected subjects to conduct the tests. The study consisted of having the user go to the progress tracking website, one at a time. On the website, all participants would attempt to accomplish the same three tasks. The first task was to look at the website interface and write down where would they click if they were going to add a student to the school, to create a new course at the school, to view the progress of a student, to create a new learning activity, or to enter a test score result for an existing student. The second task was to add the new student to the "Calculus 1" course with course id "Calc1." The third task was to change the goal of the "Calculus 1" class from trigonometry to integration.

The testing took place in a computing center at a University. This location was very comfortable for the study to take place in since it is a computer laboratory opened to students, staff, faculty, and the public.

The first step into conducting the actual experiment was selecting the participants. Because studies have shown that cultural differences may have an impact on the usability study [12], the participants for this study will be a random sample of seven people at a University. This means it could mean anybody: students, staff, or faculty. The group of subjects chosen agreed to take part of this study by signing an Informed Consent form in which the team gained permission to test on human subjects according to the university's Institutional Review Board. After the approval, the users were handed pre-surveys that allowed experimenters to learn about background information on student learning applications. This pre-survey also collected demographic data on each of the subjects. After the pre-survey was completed, the subjects were then given a specific set of three tasks that they would accomplish. These three tasks were simple enough for the users to complete in a short period of time but at the same time long enough for the users to familiarize with the website. While these tasks were being performed, the experimenter was collecting data on observations and comments the users made. The final task that the subject had to complete was a post-survey in which they were asked how they felt about the application itself. This included what they thought about the interface, the visuals, the memorability, and the control they felt they had when using the website.

The data collected by the experimenters, the pre-survey, and the post-surveys were then analyzed in order to see what were the major issues the users encountered, the time it took them to complete each task, and improvements

and suggestions that the users had towards the application. One major challenge that the experimenters faced during the study was that the users did not understand that the purpose was to record their interaction with the system. Instead, the users felt they were being graded by the experimenter on whether they were doing the task right or wrong. This lead the experimenters to collect unusable observations on comments and thoughts the users had while performing the tasks given.

4. Results

The usability study helped the team understand about the mental model of the potential users. Of the users we surveyed, only one of them had some experience with Moodle [7], while 100% of them constantly used Angel Learning [1]. All of the users felt they had some experience using educational learning software. Sixty percent of the participants were female. Table 1 displays the likes and dislikes of the participants regarding the interface design. Table 2 displays the similarities among participants in performing the first task. Participants seem to agree that the web interface is pleasant to the eye and clear to the user. The fact that users liked the design of the website and had difficulties completing tasks suggests that we have a good foundation for an intuitive interface, but wording of functions need to be given more thought. Furthermore, experimenters noted that participants apt to finish their tasks after being verbally told the tasks to be accomplished by the experimenters in the experimenters' own words.

In addition, the results unanimously agree that the dropdown AJAX technology used is not an easy-to-use way of helping users find existing items in the database. They also added that the AJAX technology made the tasks confusing. The technology was used in order to minimize user clicks and to minimize load time by avoiding having users going from page to page.

Table 1: Percent of participants that Like or Dislike specific Usability Criteria

CHARACTERISTIC	PERCENT
Colors used on the website	100%
Font used on the website	100%
Page Layout of the login page	100%
Page Layout of the home page	100%
Ease of using the website	60%
Website consistency from one screen to another	80%
Feedback messages given to users	60%
Amount of information given to the user about the input fields	80%
Amount of information users are required to remember on their own in using the website	80%
Amount of control user feels they have on the website	80%
Website designed to prevent user Error	40%
Website permits for easy reversal of actions	80%

The users enjoyed the light-blue color used on the website and considered the Lucida Sans font appealing to the eye and easy to read. The login and home page were designed with the purpose of avoiding clutter and participants liked the usability of these criteria. As for areas of improvement, the participants seemed to agree that the website is not easy to use. They disliked the way the system prevents user error and disliked the feedback messages given to users. In other words, participants did not receive enough feedback that would help them prevent errors. Then, when feedback was given to the participant, it was not communicated in a manner that all participants were able to relate to and to understand. Therefore, the language of the content used on the website does not seem to match users' mental model.

Table 2: Percent of Participates' Mental Model matching the tasks links

TASK	PERCENT
Add a student	100%
Create a new course	80%
View progress of student	60%
Create new learning activity	100%
Enter a test score	80%

In order to improve the usability of the application, the usability study suggests that the team needs to rename several of the functions in order to better match the mental model of the user.

The study also recommends that the dropdown box *AJAX* technology to be replaced by a more intuitive, easy to use technology. The *AJAX* application gave most of the users problems recognizing what they thought the tasks asked them to type on the space provided instead of instructing them to click on.

5. Conclusion

As already mentioned, this is the first usability testing that has been performed on the progress tracking application. Although there is room for improvement, the application proved to have a strong foundation for a clear and intuitive interface.

As for areas of improvement, the team learned that the application is not easy to use due to the wording of the sections. More to the point, the application seems to not match the user's mental model because 40% of the times the participants were asked to perform a task, they did not know where to click in order to execute the task in question. For example, when the experimenters asked the participants to create a new learning activity, only 60% of the participants were able to click on the correct link. Experimenters noticed that after having verbal transmission of tasks to the users, using the technical jargon, users were able to finish the tasks. As a result, in order to increase the usability of the software, interface developers will work on minimizing the technical jargon on the website. The experiment showed that technical jargon confuses and intimidates the users. In the future, the team will work on learning more about the less skilled users' vocabulary in order to help the application be easy to use to a broader audience by decreasing the learning slope in terms of language [4].

Additional to the wording of the website, the participants indicated that the system is not designed to prevent user error. In fact, one of the observations the experimenters had is that participants constantly asked questions when they made a mistake or where unsure of where to click. The idea of our student progress tracking project is that in order to be successful, we need to provide a high value of utility to the user without having the user need to read a complicated manual. More to the point, we would like the user to utilize the application without a manual at all. Having the participants ask questions during the experiment tells the developers that work needs to be done in order to make the application more intuitive.

Although the development team received valuable feedback from the study that will definitively be used for next development iterations in order to make the software match the users' mental model, the study did not provide information of what the users' mental model actually looks like. The development team would have given very good use to more specific information describing what the users expected the buttons to say. This need for more information goes along with the need of refined procedures and methods when there is a demand for research in usability [8] such as the existing

demand for research in usability of student learning related software.

Future research will continue to be performed in order to make improvements on the web application and eventually have it introduced to the market place.

6. References

[1]. ANGEL Learning Management Suite. Blackboard Home. July 2000. Web. 19 Nov. 2010. <http://www.angellearning.com>.

[2]. *Determining the importance of key criteria in web usability.* **Pearson, J. Michael.** 11, s.l. : Emerald Group Publishing Limited, 2007, Management Research News, Vol. 30, pp. 817.

[3]. *Designing the User Interface: Strategies for Effective Human/Computer Interaction*. **Shneiderman, B.** 1998, Reading, MA: Addison-Wesley.

[4]. *Human Computer Interface Design Guidelines.* Brown, C. M. Norwood, NJ: Albex, 1989, pp. 59.

[5]. Is the User Always Right? **Bath, G.** Computer Systems in Southern Africa Oct. 1983: pp. 12. Computer Database. Web. 11 Mar. 2010.

[6]. *Key usability factors of service-oriented web sites for students: an empirical study.* **Nathan, Robert J.** s.l. : Emerald Group Publishing Limited, 2008, Online Information Review, pp. 302-324.

[7]. *Moodle. Moodle.org: Open-source Community-based Tools for Learning*. Web. 19 Nov. 2010. <http://moodle.org/>.

[8]. *Re-engineering the usability-testing process for live multimedia systems.* **Georgievski, M. and Sharda, N.** 2, s.l. : Emerald Group Publishing Limited, 2006, Journal of Enterprise Information Management, Vol. 19, pp. 223-233.

[9]. *The role of pleasure in website success.* **Wulf, Kristof De.** 4, Amsterdam : Elsevier Science Publishers, 2006, Information and Management, Vol. 43.

[10]. *Usability studies of WWW sites: heuristic evaluation vs. laboratory testing.* **Laurie, Kantner and Stephanie, Rosenbaum.** New York : ACM, 1997, ACM Special Interest Group for Design of Communication.

[11]. *Web site design and content management.* **Cox, Andrew.** 3, s.l. : MCB UP Ltd, 2001, VINE, Vol. 31, pp. 3-5.

[12]. *What is part of a usability test?* **Clemmensen, Torkil.** New York : ACM, 2008, Conference on Human Factors in Computing Systems, pp. 3603-3608.

[13]. Windows 95 easy? says who? **Angus, Jeffrey Gordon**. Computerworld, 29 (34), p. 41.

Usability Evaluation for Remote Access Application: LogMeIn.com

Ayuba Audu
Florida Institute of Technology
aaudu@my.fit.edu

Joseph Noel
Florida Institute of Technology
jnoel@my.fit.edu

Abstract

A usability study was conducted to identify strengths, weaknesses, and recommendations for improvements to an online technical support tool called LogMeIn. The study was performed by six participants between the ages of 19 and 22 from a university. The website for the study was logmein.com. Each participant was given a set of instructions and a pre-survey. The usability tested the participants on setting up a computer with the LogMeIn software and then using the online software to gain access to the computer. The participants were looked over upon the experimenters throughout the process to ascertain questions if needed. A post-survey was administered with open and closed questions to establish strengths, weaknesses, and recommendations for the website. The results of the study are presented along with future recommendations.

1. Introduction

Today's world is centered completely around the growing use of technology, both at home and in the work environment. We are moving towards a scenario where one is not tied down to a single device, but multiple devices interwoven with multiple platforms that allow a user to perform a task in real time from one device to the next[5].

The use of online tools, like LogMeIn, play an important role in fixing computer related issues. Usability studies are conducted to improve the interaction between the user and website. As the scope of devices supported by any web application increases, so should the focus on usability.

A usability study was conducted to identify strengths, weaknesses, and recommendations for improvements to an online technical support tool called LogMeIn. The study was performed by six participants between the ages of 19 and 22 from the university. The website for the study was logmein.com. Each participant was given a set of instructions and a pre-survey. The usability tested the participants on setting up a computer with the LogMeIn software and then using the online software to gain access to the computer. The participants were looked over upon the experimenters throughout the process to ascertain questions if needed. A post-survey was administered with open and closed questions to establish strengths, weaknesses, and recommendations for the website. The results of the study are presented along with future recommendations. The usability study conducted is Phase 1 of a project to improve the usability of websites with technical tools.

2. Literature Review

Usability is a paramount subject in Human-Computer Interaction (HCI) since it is the characteristic that popularly refers to quality of the user interface [1]. The International Standards (ISO) sees usability as [2]: ***The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context.***

The research conducted was centered on a web application called LogMeIn. A web application allows a user to remotely initiate actions from within a browser that are processed partially by the web server in the front end with the assistance of the application and database server in the backend. "A web application is an application which has been specifically designed to be executed in a web-based environment; it is more than just a set of web pages setup with navigational links." [6]This model creates a highly scalable platform for interaction over a network.

As the scope of devices supported by any web application increases, so should the focus on usability. The key elements involved in building any web application that involves usability throughout the design process are Planning, Analysis, Mock-ups and Prototypes, Production, Testing, Launching, and Maintenance [4]. Without these seven key elements any web application will be challenging to use at best. With the current move to the cloud computing era where users access Software-as-a-Service (SaaS), web applications have really started to come to maturity. The LogMeIn platform capitalizes on this fact and tries to provide remote access to PC's, Macs and smart phone devices[3] through its web interface which allows users to perform actions on a remote device like they are physically interacting with it. The research focuses on directing the users to get the application up and running on a PC based platform and perform tasks such as connecting to the remote device and copying files via the web interface while we try to pull usability data from them.

The essential aim of all web applications is to facilitate the accomplishment of one or more tasks that the user needs to get done intuitively within the application and we try to keep this singular fact in mind as we test the usability of the LogMeIn web application[7].

3. Methodology

The usability study was conducted to increase usability literature specific to website tools used in facilitating accessing remote machines. The specific tool studied was LogMeIn Free. The research identifies the challenges that users encounter when dealing with the use of online tools to access computers located in another location. The study was composed of six steps used to setup and run the usability study on the LogMeIn website and tool.

The first step was to pick the tasks to test for our usability study. Two pre-tests were run to fix any problems in the instructions and to get feedback for improvements. The tasks consisted of a two part system to which the first task was to have the user login to the demo account to setup the software on the 'remote' computer. After the process for installing the software was complete, the second task involved the user using another computer to access the 'remote' computer through the website. The end of the second task allowed the user to perform whatever action they wish to perform on the computer. As part of the identification of the study, the following steps were determined to be the process each participant would go through to participate in the testing session:

- Reading of the pre-test instructions
- Reading and signing of the informed consent
- Reading of the task instructions
- Filling out a pre-test survey
- Performing the tasks
- Filling out the post-test survey
- Debriefing

The second step of the usability study was to gain approval, through the Institutional Review Board at the university, to conduct a test on human participants. Only after approval was given was the study able to take place.

The third step of the usability study was to prepare the documentation for the testing which consisted of the following:

- Informed Consent Form: The consent form provides a summary of the usability study and a waiver of liability that for the participant. The form informs the participant's that their participation is voluntary and that they can stop whenever they choose.
- Pre-Test Instructions: The initial instructions are general and contain information about the purpose of the study, risk of the experiment, time limit, and other notes of encouragement for the participant. There is also a brief to-do list for before, during, and after the study.
- Pre-Test Survey: The survey consisted of demographic questions regarding age, gender, college level, and ethnic background. There are also questions pertaining to computer experience, and knowledge of LogMeIn.
- Participant Written Instructions: These instructions are the tasks that the participant must perform for the experiment. These include the process of installing the software and using it.
- Experimenter Form: The form contains four columns for the experimenter to document the participant's comments and time of tasks during the test. The categories of the column are task step, task time, experimenter's comment, and customer's comment.
- Post-Test Survey: The survey contained closed ended questions regarding the website's colours, font, layout, consistency, etc. as well as a "Dislike" or "Like" rating for each question. There were also open-ended questions regarding what the participant likes the most and least about the website; as well as, what the participant thought of for improvements to the website.

The fourth step was the process of recruiting participants to participate in the usability study. The participants consisted of undergraduate and graduate students from the university. Extra credit, monetary prizes, nor gifts were given to the participants for participating in the study. All participation was strictly voluntary whether they participated or not.

The fifth step was to perform the actual study with the five subjects. Testing took place in a controlled office setting with the two experimenters and the participant. A laptop and desktop with Internet connection was provided for the participants to perform the study. The participants were randomly pulled from outside the office in a polite manner to perform the study. For confidentiality reasons, the data collected did not contain the names of the participants or any unique identifier of the participants. Furthermore, the data collected cannot be linked to any participant in the study.

The sixth step was to analyse the data collected from the study including but not limited to the experimenter form, pre-survey, and post-survey. All aspects of the study were conducted to gather information to identify recommendations for usability improvements to LogMeIn. The pre-survey enabled the researchers to gather background information on the participants to look for possible links between certain backgrounds like computer experience and age.

4. Results

The usability testing study resulted in the analysis of the data indicating many like findings using the LogMeIn website. Of the six participants 66% of them had heard of LogMeIn and 33% had not. 16.7% of the participants have some computer experience, 50% have

significant experience, and 33% have expert experience. The participants were all male and consisted of 16.7% Africa American, 66.6% African, and 16.7% Caribbean descent. All the participants liked the font, operation tools layout, ease of use, consistency, information for account setup, information for computer setup, amount of control, information needed to remember, and information to take control. 33% of the participants disagreed that the website was designed to prevent user error. It took an average of 3 minutes and 48 seconds to complete the first task and 1 minute and 49 seconds to complete the second task. On average the whole process took 5 minutes and 37 seconds to setup the computer and gain access to it.

Table 1. Pre-Survey Results

How ofter do you need information stored on a computer that is not accessible?					
Frequently	Often	Sometimes	Never		
33.30%	33.30%	33.30%	0.00%		
List any software tools you have used to access files from your remote machine?					
Remote Desktop	Team Viewer	X Windows	NX	Google Apps	None
16.60%	33.30%	16.60%	16.60%	16.60%	16.60%
What operating system do you most use?					
Windows	Apple	Linux	Other		
100.00%	0.00%	0.00%	0.00%		
What level of experience do you have with using a computer?					
Expert	Significant	Some	Little	None	
33.30%	50.00%	16.60%	0.00%	0.00%	
My age range is?					
18 or younger?	19-20	21-22	23-24	25-49	50 or older
0.00%	83.30%	16.60%	0.00%	0.00%	0.00%
I am a?					
Freshman	Sophmore	Junior	Senior	Graduate	Faculty/Staff
0.00%	0.00%	33.30%	66.66%	0.00%	0.00%
My ethnic background is?					
Caucasian	African American	Indian	Hispanic	Caribben	African
0.00%	16.60%	0.00%	0.00%	16.60%	66.66%
I am a?					
Male	Female				
100.00%	0.00%				

Table 1 contains the results of the pre-survey which includes the demographic information we collected form out participants.

Post Survey

The post-survey consisted of both close and open ended questions. The close ended questions asked the participants whether they liked or disliked the account setup, colors, consistency, ease of use, fonts, feel of control, home screen layout, amount of information to remember, tools layout, process of taking control of the remote machine and chance of user error while interacting with the application.

Table 2. Post-Survey Results

Colors	***Account Setup***
Like	Like
100.00%	100.00%
Font	***Computer Setup***
Like	Like
83.34%	100.00%
Homescreen Layout	***Taking Control***
Like	Like
83.34%	100.00%
Operating Tools Layout	***Information to Remember***
Like	Like
100.00%	100.00%
Ease of Use	***Feel of Control***
Like	Like
100.00%	100.00%
Consistency	***User Error***
Like	Like
10.00%	66.66%
Feedback	***Ease of Undoing***
Like	Like
83.33%	83.33%

Table 2 contains the post-survey close-ended questions results with the "like" and "dislike" responses for various usability aspects of the LogMeIn web application.

The open ended questions gave participants the opportunity to write their own comments on what they liked best and least about the LogMeIn web application, and their recommendations for improvement.

Most of the participants were satisfied with web application as we saw a lot of positive comments complimenting the ease of installation and use, the lack of limitations with what you can do with the remote machine and the feel of control of all that the web application did while one was using it. The negative comments were sparse at

best, but we still got a valuable piece of information from the participant who mentioned the screen resolution being low and in some areas the font and icons being small.

The comments pertaining to improvement only concerned the issue of the resolution being low resolution, and fonts and icons being small. The question that asked for recommendation for improvement was a surprise for the participants, since they felt that without any restrictions on the actions they could perform on the remote machine and the comforting sense of control they felt when using the web application lacked any easily apparent need for improvement.

	Table 3. Data Collection (Seconds)		
	Average Time Taken Per Task		
Task 1.1	247		
Task 1.2	118		
Task 1.3	282		
Task 1.4	139		
Task 1.5	165		
Task 1.6	179		
Task 1.7	237		
Task One Average	228		3 Minutes 48 Seconds
Task 2.1	176		
Task 2.2	54		
Task 2.3	103		
Task 2.4	192		
Task 2.5	59		
Task 2.6	104		
Task Two Average	109		1 Minutes 49 Seconds
Overall Average	337		5 Minutes 37 Seconds

Table 3 contains the timed results of the individual tasks performed by the participant. Table 3 contains the computed averages of task 1, task 2, and overage completion of the tasks.

The results suggested that the usability of the website was strong, allowing the participants to easily move through the tasks. The participants had a strong control over the system. The participants noted that the instructions were easy to follow and moving through the website was smooth. Future improvements were noted to include a file transfer option between computers.

5. Conclusion

The LogMeIn website has strong usability components that allow users to perform the tasks they need to with little problems which we can summarize under the following headings;

1) Easy to learn: All the participants with zero experience using the LogMeIn web application were able to follow on screen feedback coupled with the simple instruction sheet that was provided to become adept with the web application to accomplish tasks assigned.
2) Useful: All the participants commented that they could easily see this application being as common as the flash drives they carry around as it allows them full control of the remote machine to perform simple tasks such as printing and emailing an important file to yourself to the more complex helping someone else use your remote machine.
3) Easy to use: The application required no great labour on the participants end to use. The feedback was instantaneous, the important information was located clearly and organized in such a way that it does not obstruct one as they try to perform a task on the remote machine that they connect to.
4) Pleasant to use: The singular act of being connected to the remote machine that you own and recognize and being able to use it with virtual no restrictions in the manner you are used to was very appealing and pleasing to the participants. It had a very original feel to it.

There is still a lot of room for improvement as noted in the comments from the participants,

although the overall satisfaction of the website is good.

Future research suggestions would be to include users with different experience levels, people, and tasks. With varying the experience levels of the participants, we would seek to obtain quality related information on how a wide range of people from novice to expert respond the web application and how their general experience with computers affects their ability to use LogMeIn.

With varying the people we would see to obtain quality related information the possible differences in the way different genders and cultures view LogMeIn, since as Al Hunaiyyan argues, "culture is a discernible variable in the acceptance of an interface. [8]"

With varying the tasks that the participants are assigned to perform, we would be seeking to obtain quality related information on how task complexity affects the participants view of LogMeIn and also whether any features of LogMeIn hinder the participants from performing the tasks more easily.

We strongly feel that if the information obtained from such improvements were to be implemented in future revisions of the web application, it would be easy to imagine a situation where anyone would be able to use it with ease.

6. References

[1] Parlangeli, O., Marchingiani, E. & Bagnara, S.,"Multimedia in Distance Education: Effects of Usability on Learning", Interacting with Computers, 1999, pp. 12(1): 37-49.

[2] International Standards Organization (ISO), "ISO-9241: Guidance on Usability Standards",http://www.iso.ch/iso/en/CatalogueListPage.CatalogueList?I, 1998,CS1=13&ICS2=180

[3] About Us page LogMeIn, retrieved December 9th, 2009 from https://secure.logmein.com/US/about/aboutus.aspx,

[4] Darren Gergle, Tom Brinck, Scott Wood, "Practical Usability Methods in Website Design", Conference on Human Factors in Computing Systems, ACM, New York, 1999, pp. 110-111

[5]Martin Gaedke, Johannes Meinecke, Andreas Heil, "FDX – Federating Devices and Web Applications", Proceeding ICWE '06 Proceedings of the 6th international conference on Web engineering, ACM New York, NY, USA, 2006, pp 95 - 102

[6]Vince Bruno, AudreyTam, James Thom, "Characteristics of Web Applications that Affect Usability", Proceedings of the 17th Australia conference on Computer-Human Interaction, Computer-Human Interaction Special Interest Group (CHISIG) of Australia, Australia, 2005, pp. 1-4

[7]Marilyn Hughes Blackmon, Peter G. Polson, Muneo Kitajima, Clayton Lewis, "Cognitive Walkthrough for the Web", Conference on Human Factors in Computing Systems Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, New York, NY, USA, 2002,pp 463 - 470

[8]Panagiotis Zaharia, Anthony Papargyris, "The gamer experience: Investigating relationships between culture and usability in massively multiplayer online games**",** ACM Computers in Entertainment, Vol. 7, No. 2, Article 26, ACM New York, NY, USA, 2009.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Heuristic Evaluation of State Electronic Government to Promote Usability for Citizens of All Ages

S. Ann Becker, Deborah Sater Carstens and Tamika Maree Linton
Florida Institute of Technology
abecker@fit.edu, carstens@fit.edu, tlinton2008@my.fit.edu

Abstract

A heuristic evaluation was conducted to identify potential usability barriers associated with fifty state government Web sites by citizens of all ages. The evaluation criteria in this study included readability, navigation, consistency, reading complexity, understandability, and the availability of help information. The heuristic evaluation took into account normal aging factors associated with older adults, sixty years plus. The heuristic evaluation also took into account readability from the perspective of reading comprehension requirements for adults of all ages. The data findings revealed potential usability issues that could impede the use of state site. These results are discussed in terms of their practical application for improving state government Web sites. Web developers of State government sites may find these results insightful in promoting site usability for all citizens.

1. Introduction

A study was conducted to identify potential usability barriers of state government Web sites that offer online resources to U.S. citizens of all adult ages. This study assessed electronic government (e-government), at the state level, by applying usability heuristics as general principles for quality Web site design. For this study, heuristics are considered rules of thumb in pursuit of designing universally accessible Web sites. The practical application of such heuristics would uncover potential usability impediments to be addressed during Web design activities.

Universal accessibility focuses on design efforts such that a Web site is usable by all adult age groups. Usable Web designs, in general, take into account design layout, consistency across pages, navigational schema, and help services, among others [20, 21 & 22]. Usable Web designs, particularly for older adults sixty years plus, account for degrading vision and impaired motor skills associated with normal aging. Usable Web designs should address reading comprehension of Web contents in ensuring that all citizens have access to information content closely aligned with reading skill levels.

The heuristics that were applied are loosely defined as: consistency within and across Web pages; ease of use associated with navigational schemas; readability associated with design layout, font style, size, and type; reading complexity associated with information content and online help support. These heuristics are considered popular in industry and academia for evaluating usability of Web sites when taking into account the needs of targeted users.

This study customizes the heuristics to account for a user profile (e.g., vision associated with normal aging) and the objective of using the Web site (e.g., informational content). This customization is done to gather meaningful data about usability associated with a particular Web

site. Bainbridge [2] and Carstens et al. [7], for example, tested Web sites for travel sales using specific heuristics related to online accommodations. Becker [3] assessed the usability of health-related Web sites by older adults in search of health resources. Carstens and Becker [8] conducted a usability study on state Web sites that offer resources on government performance.

The heuristics evaluated the usability of a state e-government site by adults and older adults sixty years plus. The study was performed via the application of system evaluations [13] whereby a Web interface was used to measure various dialogue elements against the heuristics. The results of the heuristic evaluation included both impediments and strengths of Web site usability. The results are summarized with the objective of improving usability aspects of state e-government. Future research proposes to broaden the scope of this study in pursuit of universal accessibility of state e-government sites by all citizens.

2. Literature Review

Researchers and practitioners alike have studied usability in order to develop Web sites that are navigable, consistent, appealing, clear, simple, and forgiving of user mistakes [6 & 17]. Existing user interface design recommendations were extended to include user interfaces for the Web [14, 27, & 28]. Those experienced in designing user interfaces provided heuristics and guidelines for designing Web pages, often by identifying design layout, navigation, and performance issues associated with particular Web sites [11 & 29]. Jakob Nielsen has been instrumental in providing heuristic guidelines for usable Web design [20, 22 &23].

Web usability research, in terms of universal accessibility, has been extended to include Web designs that are usable by older adults [3 & 5]. The older adult population in the U.S. is one of the largest groups of users benefiting from online access to government resources. Since 1990, the number of Americans sixty-five or older has more than tripled. Older adults will comprise about 20% of the total population by 2030, which is twice more than the percentage in 2000 [1]. Vision, motor skills, and literacy play a role in usable Web designs for older adults.

A Web usability impediment for older adults, for example, may include a small font size (e.g., 8 to 10 point font) for informational links. Color usage, such as pastel or saturated foreground and background combinations, may make it difficult to read page content [4]. Web sites utilizing mouseover technology often require precision movement to navigate the site, which may be difficult for those with lost sensitivity in fingertips, stiff joints, or shaky hands [9, 32, & 33].

The U.S. National Institute on Aging (NIA), in conjunction with the U.S. National Library of Medicine (NLM), has published guidelines for designing senior-friendly sites that are based on scientific findings from research in aging, cognition, and human factors [15 & 24]. These guidelines include the following: sans serif typeface, 12 point or greater font size, mixed case letters in the text body, left justification, plan background image, and text effects used only in headlines. Several of these Web design guidelines were applied during the heuristic evaluation of state e-government sites.

Web designs may impose usability impediments when reading comprehension levels are misaligned with reading grade levels representative of the general public. The heuristic evaluation applied the Flesch Reading Ease readability scoring system [12] to state e-government sites to assess reading complexity.

From a universal accessibility perspective, Web content should accommodate a reading grade level having the broadest citizen reach. The National Work Group on Literacy and Health [18] recommends a fifth reading grade level for all adults. The University of Utah's Health Sciences Center [30] states that an average adult typically reads one to two grade levels below their last grade completed; thus, it recommends a sixth-grade reading level.

Wickens and Holland [31] identify age as having an impact on literacy such that comprehension of written material progressively declines. It is recommend by Doak et al. [10] that content be written at a fifth grade reading level to accommodate older adults.

3. Methodology

A study was conducted to assess the usability of each of the fifty state e-government sites. The usability study consisted of the following steps in planning, conducting, and assessing the selected Web sites.

Step 1: *Develop usability criteria to assess each state Web site in terms of potential usability barriers*. The assessment criteria focused on usability by citizens of all adult age groups. The criteria consisted of text resizing capabilities, readability, navigation, consistency, reading complexity, and availability of help information.

Step 2: *Evaluate each state Web site utilizing the usability criteria.* Tables 1 through 4 display the results of the evaluation consisting of strengths and weaknesses. Figures 1 through 5 are illustrative of assessment findings.

Readability	# of Sites	%
Less than 12 point font size	22	43
Text resizing does not work	19	37
Images overlaid with text	18	35
Use of serif font	7	14

Step 3: *Identify potential usability barrier.* The usability issues for each Web site are summarized based on the application of Step 1 criteria.

Step 4: *Summarize the results for promoting universal accessibility of state e-government sites.* The conclusion section discusses usability improvements in pursuit of universally accessible state e-government.

The heuristic evaluation was conducted by applying widely-accepted Web guidelines [22 & 23], NIA guidelines for senior-friendly Web sites [24], and the Flesch Reading Ease score [12]. The heuristic evaluation used Internet Explorer version 7.0 to display Web pages. The color monitor size was 14 inches and Internet connectivity was broadband with a speed of 54 Mbps (megabytes per second). All Web pages were accessed in late 2009 and early 2010 with no visible changes to the Web sites during the evaluation process.

4. Results

The results of the usability assessment of the fifty states e-government sites are presented for each criterion previously identified.

Readability issues with Web content. The evaluation results identified design components that may adversely impact readability. These components consisted of: (1) text format being smaller than 12 point font size; (2) resizing text option distorting Web site graphics and text; (3) images overlaid text; and (4) use of non-sans serif text format.

Table 1: Usability Results for Readability[i]

Table 1 displays the results associated with each readability issue assessed. Over 40% of the Web sites used a font size less than 12 point for

some of its content. For one of the state Web sites, for example, content was presented in a small font size (7.5), which couldn't be resized using the browser resizing feature.

Over one third of the sites failed to fully support text resizing using the browser's resizing feature; or when text resized, it overlaid images. Figure 1 displays an example from a state government Web site of content that becomes distorted when the page is resized.

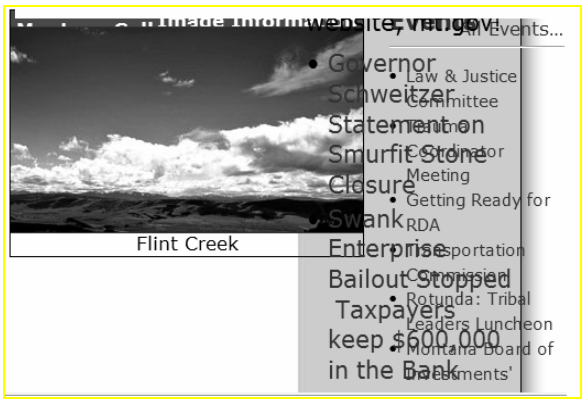


Figure 1: Text Resizing [16]

Table 1 also identifies about 14% of the sites used a serif font type. About 86% complied with the use of sans serif font type.

Reading complexity issues with Web content. The evaluation results identified design components that may adversely impact reading comprehension particularly for those with vision impairments or degradation due to aging. Only those Web sites with Medicare information had the Flesch Reading Ease score [12] calculated.

The Flesch Reading Ease score rates text on a 100-point scale based on the average number of syllables per word and words per sentence [12]. The higher the Flesch Reading Ease score, the easier it is to understand the document. For most standard documents, recommendations are for a Flesch Reading Ease score of approximately 60 to 70 (reading grade level of 6th -7th grade).

Table 2 displays the number of Web sites that fell within each category of the Flesch Reading Ease score, which is mapped to the corresponding readability level.

Table 2: Readability Score

Flesch Reading Ease Score	Readability Level	Number of Websites
0 - 29	Very difficult	7
30 - 49	Difficult	20
50 - 59	Fairly difficult	9
60 - 69	Standard	1
70 - 79	Fairly easy	0
80 - 89	Easy	0
90 - 100	Very easy	0
TOTAL		37

It is noted that 98% of the evaluated sites would require a comprehension level higher than seventh grade. The results show that almost a fifth of the sites had content considered to be very difficult in terms of reading comprehension level.

Reading complexity is further illustrated by the following sentence appearing on the New Hampshire state government site under "Seniors", "Diseases and Conditions", and "Ashma" [19]. The sentence has 48 words with a Flesch Reading ease score of 0 (as calculated by Microsoft Word.)

"*The NH State Asthma Plan, developed by an advisory council representing stakeholders from across New Hampshire, describes the challenge asthma presents in the state and makes recommendations for action in the areas of clinical services and disease management, indoor and outdoor environments, public awareness and education, and surveillance*."

Figure 2 is an example of a Web site that displays images overlaid with text. The homepage shows several examples of text

overlaying images. For aging vision, in particular, text overlay can impede readability.

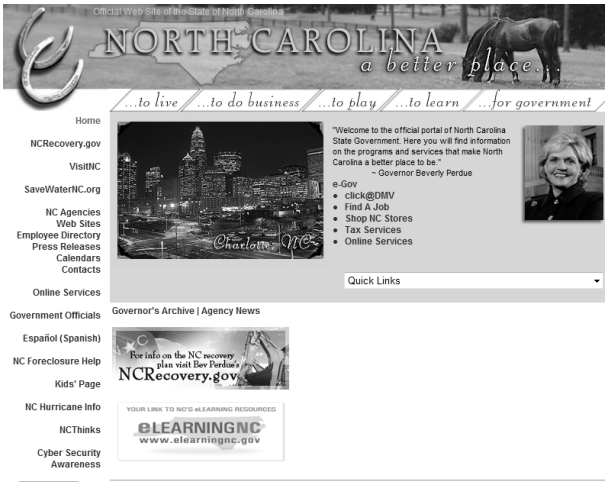


Figure 2: Images overlaid with text [25]

Navigation is not always intuitive. There are several design components related to navigation in determining if a Web site is intuitive for a user. Design components that adversely impact the use of the Web site: (1) Breadcrumb trail are is not used for navigational support; (2) link or menu option directs the user to an external Web source without notification about leaving the e-government site; (3) table of contents feature does not exist to assist the user in navigating throughout a site; (4) images without any form of labeling are used to provide navigation to users; and (5) users click through more than three links to access desired information. Table 3 captures these components and displays the results associated with each navigation issue assessed on the 50 states Web sites.

Table 3: Usability Results for Navigation

Navigation	# of Sites	%
Breadcrumb trails not used	23	45
Links lead to external site	13	25
Table of Contents not used	10	20
Images (with no labels) are used to navigate site	20	39
More than 3 clicks to access information	32	63

The table shows that 45% of the Web sites did not use breadcrumb trails and about 20% did not use a table of contents feature. Almost 40% used images for navigational purposes without providing content that identified them as such. For navigational depth, over 60% of other Web sites required three or more mouse clicks to reach a targeted destination.

Web design does not provide supporting information. There are several design components that may impact supportability for a user: (1) lack of contact information on the Web site; (2) no live online help (e.g., online chat option); (3) no feedback option for users to provide comments regarding Web site usability; (4) no privacy policy; and (5) no accessibility policy identifying how users with disabilities can access the web site thereby ensuring equal access to electronic and information technologies.

Table 4 summarizes data findings as they relate to supportability. For several sites, a privacy policy could not be found. Approximately 14% of the sites did not have an accessibility policy. About 40% of the Web sites did not provide online assistance. Approximately 20% of the Web sites did not have contact information readily accessible and one-third did not have an onsite feedback option.

Table 4: Usability Results for Supportability

Supportability	# of Sites	%
No contact information	10	20
Site has no live online help	20	39
Site has no feedback option	17	33
Site has no privacy policy	2	4
Site has no accessibility policy	7	14

Figure 3 provides an example of a Web site that does provide alternate text labeling on an image thus enabling individuals with disabilities

requiring screen readers to be able to navigate the site successfully. Screen readers can accurately identify the text label associated with an image when alternate text labeling is utilized. Without alternate text labeling, the screen reader will only be able to say "edit". This fails to provide supporting information to users with disabilities in navigating a site.

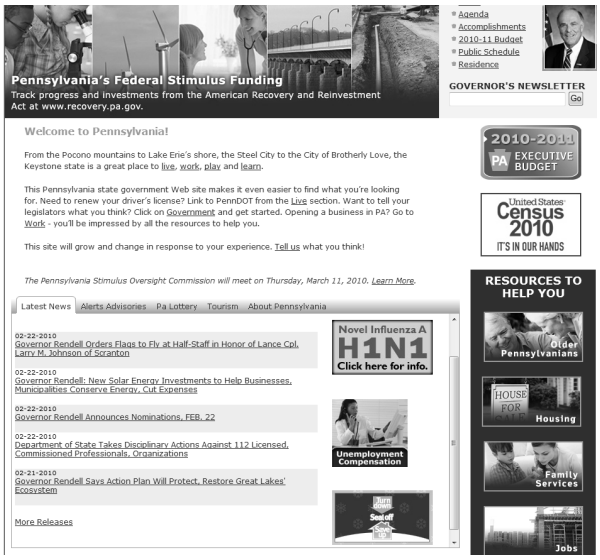


Figure 3: Alternate Text Labeling on Governor Rendell's Image [26]

5. Conclusion

The objective of evaluating state e-government sites was to uncover potential usability barriers for which design improvements can be made. It was found that usability issues exist on e-government sites inclusive of those associated with normal aging and reading complexity.

The following discussion summarizes these findings and highlights design improvements associated with widely-accepted design guidelines, NIA guidelines for senior-friendly sites, and literacy recommendations.

Readability - The results of the heuristic evaluation showed that e-government sites could readily be improved, in terms of usability, by supporting a text resizing feature and using a sufficiently large font size in sans serif font type. A Web site that supports the resizing of text promotes readability by any user with vision impairment, due to normal aging or otherwise. The use of 12 point font size or larger and the use of sans serif font type (e.g., Arial, Helvetica) support readability particularly for those with aging vision. Images should not be overlaid with text, as it obstructs readability of information content.

Reading Complexity. The results of the heuristic evaluation showed that e-government sites need to be sensitive to the reading grade level associated with comprehending information content. This is particularly important when providing resources to senior citizens given that many do not read at higher grade levels. The Flesch Reading Ease score should not exceed the "standard" level to ensure greater readability and ultimately to support universal accessibility by all citizens.

Navigation - The results of the heuristic evaluation showed that e-government should improve navigation schema for a more intuitive design. Navigation design improvements are particularly important for less computer-literate and older adults taking into account the impact of aging on cognition. A navigation feature, referred to as a breadcrumb trail, should be used to easily identify user location within a site. A common bread crumb trail design is to show at the top of the site traversed links such that a user can return to visited pages. Links that lead to an external site should be properly labeled (or a message displayed) such that the user is aware he or she is leaving the site. Each site should have a table of contents to easily guide the user to the location of desired content. All images need to be properly labeled for improved readability and to support universal access by those with vision impairments or degradation due to normal aging. Commonly used design guidelines recommend information content being found in three or fewer clicks. This is

particularly important for older adult users in accounting for normal aging and its impact on cognition.

Supportability – The results of the heuristics evaluation identified supportability issues that need to be addressed. E-government sites should clearly display a link to contact information consisting of a phone number, email address and/or mailing address. E-government sits having online help features provide real-time support for users in search of government resources. Given that e-government sites are to support all U.S. citizens, it would be important to provide a feedback mechanism for design improvements. Privacy and accessibility policies have become a widely-accepted design norm. All e-government sites should have these policies readily accessible in meeting the information needs of state constituencies. An accessibility policy, in particular, should be part of all e-government sites in providing universal access to all users including those with disabilities.

The findings suggest that usability improvements are needed by e-government sites in promoting universal access by adults of any age including seniors. Future research will expand this study to include a more comprehensive review of design components. Further emphasis will be place on reading complexity of information content associated with government performance.

6. References

[1] Administration on Aging. (2002). A profile of older Americans: 2002. Administration on Aging, U.S. Department of Health and Human Services. Retrieved April 26, 2009, from www.aoa.gov/prof/Statistics/profile/2002profile.pdf.

[2] Bainbridge, A. (2003, February). Hotel booking process: Design & usability. (1.02). Retrieved January 25, 2009, from www.travelucd.com/research/pdf/hotel_booking_process_february2003.pdf.

[3] Becker, S. A. (2004). A Study of Web Usability for Older Adults Seeking Online Health Resources. *ACM Transactions on Human-Computer Interaction*, *11*(4), 387-406.

[4] Becker, S. A. (2004). E-government visual accessibility for older adult users. *Social Science Computer Review*, *22*(1), 11-23.

[5] Becker, S. A. (2009). Web accessibility and critical issues facing older adult users. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology,* 2nd Ed., Hershey, PA: IGI Global, 4041-4046.

[6] Carstens, D. S. (2005). Cultural barriers to human-computer interaction. In S. Marshall, W. Taylor, & X. Yu (Eds.), *Encyclopedia of Developing Regional Communities with Information and Communication Technology*. Niagara Falls, New York: Idea Group Inc.

[7] Carstens, D. S., Feizi, S. & Malone, L. (2009) Traveling Abroad? A Usability Assessment of Travel Sales Web sites, Journal of Management & Engineering Integration, 2(1), 17-24.

[8] Carstens, D., and Becker, S. A. (2009). A Heuristic Study on the Usability of State Government Performance Data Web Sites. *Issues in Innovation Journal,* 4(1).

[9] Chaparro, A., Bohan, M., Fernandez, J. E., Choi, S. D., AND Kattel, B. (1999). The impact of age on computer input device use: Psychophysical and physiological measures. *International Journal of Industrial Ergonomics, 24*, 503-513.

[10] Doak, C., Doak, L., and Root, J. (1996). *Teaching Patients with Low Literacy Skills*. 2nd Edition, J.B. Lippincott Company, Philadelphia, PA.

[11] Flanders, V., & Willis, M. (1998). *Web Pages That Suck.* San Francisco, CA: SYBEX.

[12] Flesch Reading Ease Readiability Score. (2010). RFP Evaluation Centers. Retrieved on November 20, 2010, from http://rfptemplates.technologyevaluation.com/Readability-Scores/Flesch-Reading-Ease-Readability-Score.html.

[13] Fichter, D. (2001). Testing the Web site usability waters. *Online,* 25(2), 78-80.

[14] Lynch, P.L., & Horton, S. (1999). *Web Style Guide: Basic Design Principles for Creating Web Sites.* New Haven, CT: Yale University Press.

[15] Morrell, R. W., Dailey, S. R., Feldman, C., Mayhorn, C. B., Echt, K. V., Holt, B. J., & Podany, K. I. (2004). Older adults and information technology: A compendium of scientific research and Web site accessibility guidelines. National Institute on Aging, Bethesda, MD.

[16] Montana State Government Web site. Retrieved 2009 from, http://mt.gov/.

[17] Murray, G., & Costanzo, T. (1999). Usability and the Web: An overview. *Network Notes*, *61*. Information Technology Services, National Library of Canada. Retrieved October 1, 2008, from http://www.nlc-bnc.ca/9/1/p1-260-e.html.

[18] National Work Group on Literacy and Health. (1998). Communicating with patients who have limited literacy skills. *The Journal of Family Practice*, *46*, 168-175.

[19] New Hampshire State Government Web Site. Retrieved November 21, 2010, from www.dhhs.nh.gov/foryou/seniors.htm.

[20] Nielson, J. (1993). *Usability Engineering*. San Diego, CA: Morgan Kaufmann.

[21] Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods*, New York, NY: John Wiley & Sons.

[22] Nielsen, J. (1999). *Designing Web usability: The art of simplicity.* Indianapolis, IN: New Riders Publishing.

[23] Nielsen, J., & Tahir, M. (2002). *Homepage usability 50 Web sites deconstructed.* Indianapolis, IN: New Riders Publishing.

[24] NIA & NLM. (2002). Making your Web site senior-friendly: A checklist. National Institute on Aging and National Library of Medicine. www.nlm.nih.gov/pubs/checklist.pdf.

[25] North Carolina State Government Web site. Retrieved 2009 from, http://www.ncgov.com/.

[26] Pennsylvania State Government Web site. Retrieved 2009 from, http://www.state.pa.us/.

[27] Schneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Boston, MA: Addison-Wesley.

[28] Schneiderman, B., & Plaisant, A. (2005). *Designing the User Interface*, New York, NY: Pearson Addison Wesley

[29] Spool, J., Scanlon, T., Schroeder, W., Snyder, C., & DeAngelo, T. (1999). *Web site usability: A designer's guide.* San Francisco, CA: Morgan Kaufman.

[30] University of Utah Health Sciences Center. (1997). *Patient Education Materials An Author's Guide.* Office of Patient Education, University of Utah Health Sciences Center.

[31] Wickens, C. D., and Holland, J. G. (1999). *Engineering Psychology and Human Performance, 3rd Edition*, Prentice Hall, Upper Saddle River, NJ.

[32] Walker, N. Millians, J., AND Worden, A. (1996). Mouse accelerations and performance of older computer users. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting,* Santa Monica, CA, 151-154.

[33] Worden, A., Walker, N, Bharat, K., and Hudson, S. (1997). Making computers easier for older adults to use: Area cursors and sticky icons. *Proceedings of the SIGCHI conference on Human factors in computing systems,* Atlanta, GA, 266-271.

[i] Table percentages, in this paper, are rounded.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

A Novel Design for the Proximal Interphalangeal Joint Endoprosthesis

Jonathan Desa, Ha Van Vo and R. Radharamanan
Mercer University
desa_j@mercer.edu; vo_hv@mercer.edu; radharaman_r@mercer.edu

Abstract

This paper presents a noval design for the proximal interphalangeal joint endoprosthesis for arthritic patients. The materials used in this study are Ultra High Molecular Weight Polyethylene (UHMWPE) and titanium alloy (Ti-6Al-4V). The 3-dimensional finite element model of the implant was constructed and tested using SolidWork 2008. Based on the proposed size and the wall thickness of the titanium alloy for the cylindrical head of the implant, the tensile and compressive test Hoop stress values from simulation runs indicated that the chosen material (Ti-6Al-4V) is acceptable and can be used for designing the cylindrical head. Based on the yield strength of UHMWPE which is lower than the Hoop stresses predicted through simulation runs, the articulating cup designed using this material will deform permanently in the plastic region. Hence, further improvement and different material usage in designing the articulating cup were recommended. However, the proposed total joint replacement will help the patient to regain the full range motion and release the pain in the proximal interphalangeal joint.

1. Introduction

The focus of the study will be on the Proximal Interphalangeal (PIP) joint as seen in Figure 1. The PIP joint can be modeled as a hinge-joint. The volar ligament and two collateral ligaments stabilize the PIP joint [1] [2]. This setup is similar in all five phalanges except for the thumb since it does not have the PIP joint.

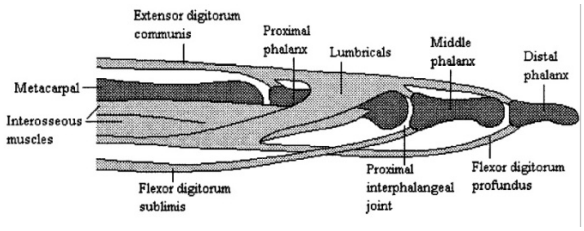


Figure 1. Cross section of the PIP joint [1]

The PIP joint movement can be described as having flexion and extension movement with rotation that is not active. This means that any rotation that the PIP joint experiences is due to the structure that the hand is trying to manipulate which may cause the joint to experience rotation. However, flexion and extension ranges from 0-100^{o} of movement from the central axis [1]. Mechanically, the joint is described as a uni-axial hinge joint with one degree of freedom [3].

In this paper, a 3-dimensional finite element model of proximal interphalangeal joint endoprosthesis was designed and tested using the software SolidWorks 2008. The proposed design uses UHMWPE for the articulating cup and titanium alloy (Ti-6Al-4V) for the cylindrical head. This improved design when properly tested, fabricated, and implanted will not only help the patient to regain the full range motion (flexion and extension) but also release the pain in the interphalangeal joint.

2. PIP joint prosthesis

Early joint prosthesis used to restore joint biomechanics in the hand began with the popular Swanson prosthesis as shown in Figure 2. This joint prosthesis has shown to relieve pain, correct range of motion and deformity. However, the benefits of this prosthesis did not last long due to implant deterioration. The Swanson implant is constructed of silicone parts and placed into the destroyed joint capsule as spacers between the phalanges where it resides in the fibrous joint capsule which provide the main fixation [4]. Due to bone re-growth, prosthesis movement, and auto-immune reactions, the silicone deteriorates over time [4].

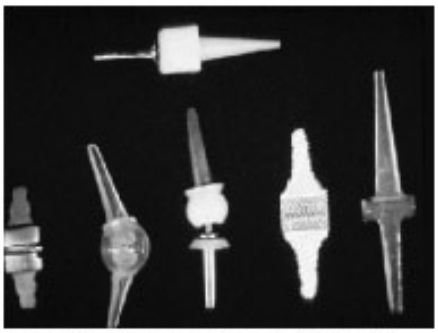


Figure 2. Swanson implants [5]

Improvements to the Swanson implant were created as time progressed and biomechanics become a more developed field. Newer designs consisted of using better materials and changing the articulating joint heads. For example, some competitors anchored the ends of the Swanson like spacers into the phalanges with metal screws. Others changed the articulating joint heads to better mimic natural PIP joint articulation. While some used new materials such as pyrolitic carbon [4]. Overall, newer designs of the PIP joint more or less followed the design criteria identified by Ash and Unsworth: mimic the natural PIP biomechanics; use bio-compatible materials that are wear resistant; fixation of prosthesis; manufacturing; surgical and patient considerations [1].

3. Methodology

The proposed design for the PIP joint is based on the design criteria identified by Ash and Unsworth [1]; it also prevents joint dislocation and minimizes destruction of surrounding structures during implantation.

The new design will consist of two components: a full Ti-6Al-4V cylindrical head and fixated rod will consist of the articulating cylindrical head of the implant; and the second component will be a cylindrical UHMWPE socket that is joined to the other Ti-6Al-4V fixated rod that will be inserted into the proximal phalanges end. The use of the ball and socket like design satisfies the criteria of preventing joint dislocation. The titanium alloy, Ti-6Al-4V was chosen due to its good material properties and its previous use and success obtained in other joint implants such as hip and knee. UHMWPE was chosen to be the articulating cup due to its favorable material properties over silicone that was used in the Swanson model. In summary, the proposed design uses a titanium alloy for the articulating head with a UHMWPE socket, which is an improvement over the old silicone design.

To further mimic the joint biomechanics, the proposed design will have a gap to allow for 100° of flexion and extension. In addition the Ti-6Al-4V stems will have ridges where it will anchor into the bone. The stems will have a thin layer of beaded hydroxyapatite formula to promote bone re-growth after implantation. The hydroxyapatite and metal anchors will be the mode of fixation of the implant into the phalanges.

Surgical procedures to implant the proposed PIP design will consist of making a vertical cut to the medial side of the PIP joint. Once the cut is made, the distal and proximal phalanges can be bent to allow exposure of the articulating surfaces. It is here that the articulating cylindrical head component will be drilled and implanted into the distal phalanges. The socket

component will be implanted into the proximal phalange. The joint will be complete once the articulating cylindrical head is placed inside the cylindrical socket. The surgeon must test to make sure the flexion and extension motion is not hindered by the implant and that fixation is secure. The surgeon will have to make adjustments as needed. After insertion, fixation, and testing, the surgeon will reattach ligaments and tendons, and suture as needed. This surgical method is different from previous implants. Usually the surgery consists of making a cut from the proximal to distal phalanges destroying the joint capsule. The surgery method to implant the proposed design destroys less surrounding structure that may further stabilize the joint prosthesis.

If the implant design proves successful, manufacturing of the PIP joint can be scaled to the other phalanges. The model developed in this study is for the index finger. Based on the measurements of the index finger, the design can be scaled to other phalanges as needed [1].

3.1. Design analysis

The choice of cylindrical head diameter and thickness of the UHMWPE cup were restricted by using the dimensions from normal PIP joints and previous implant designs [1]: (1) cylindrical cup and articulating cylindrical head has a combined length of 11 mm; and (2) the cylindrical cup cannot have a diameter larger than 11 mm [6].

The following theories and methods were used to analyze the potential design. Hertz contact stress [7] is given by:

$$b = \sqrt{\frac{2F}{\pi l}\frac{\left(1-\nu_1^2\right)/E_1 + \left(1-\nu_2^2\right)/E_2}{1/d_1 + 1/d_2}} \qquad (1)$$

where: b = the distance of the half width peak pressure where the contact stress occurs; F = the applied force from the cylinder to the cup; v_1, E_1= the material properties of the UHMWPE; v_2, E_2 = the material properties of the Ti-6Al-4V; d_1 = diameter of the UHMWPE cylindrical cup; and d_2 = diameter of the cylindrical head.

The maximum pressure at the half width distance [7] is given by:

$$p_{\max} = 2F/\pi bl \qquad (2)$$

where: $p_{\max}$ = maximum pressure applied at the half width distance, b; f = the applied force from the cylinder to the cup; b = the distance of the half width peak pressure where the contact stress occurs; and l = the length of the cylinder.

The average pressure of the contact area [8] is given by:

$$p_{avg} = (\pi/4)P_{\max} \qquad (3)$$

where: Pavg = average pressure of the contact area; and Pmax = maximum pressure applied at the half width distance.

The Hoop stress [9] is given by:

$$\sigma_\theta = P.r/t \qquad (4)$$

where: σ_θ = hoop stress; P = the internal pressure; r = the radius of the cylinder; and t = the wall thickness.

The solid works modeling will help to plot and analyze the following: von Mises stress distribution; deformation/displacement; and strain on UHMWPE socket and titanium alloy cylindrical head.

Applying the material properties of the Ti-6Al-4V and UHMWPE to the equations above, the maximum pressures were calculated for several cases. The applied force used was taken from a study conducted by Fowler and Nicol [3]. In their study, they analyzed the loads of daily activity that the PIP and DIP (Dorsal Interphalangeal) would experience. Based on their study, the maximum force was chosen out

of all activities to be the resultant applied force of 34.29N.

In this study 34.29N force was in tension and in compression. With the applied force, several cases were evaluated to determine the optimal ball diameter to inner cylindrical cup diameter based on the maximum cup diameter constraint of 11 mm. The ideal situation would have the largest socket wall thickness to minimize deformation. In addition, the larger ball size would prevent dislocation from the UHMWPE socket. Lastly, the ideal situation would provide less contact stress in both tension and compression cases. Based on initial optimization studies using different head diameters and socket thickness sizes, the case with the cylindrical head diameter of 5 mm, socket of 6 mm, and UHMWPE wall thickness of 2.5 mm was chosen.

Hoop stress was used in analyzing the tension case. This calculation required the maximum pressure of the tension case, the optimized socket thickness, and the internal diameter of the socket. Based on the criteria, the calculation yielded a Hoop stress of 2.394e+07 Pa for the tension case and 1.1970 e+07 Pa for the compression case. The yield strength of the UHMWPE is 2.1e+07 Pa and therefore, deformation will occur for the tension case.

SolidWorks 2008 was used to do stress, displacement, and strain studies to see how the contact stress from loading the PIP joint with 34.287 N in compression and tension would affect the UHMWPE cup as well as the articulating cylindrical head. In addition, SolidWorks was used to confirm the Hoop stress calculations suggesting that deformation will occur. The compression and tension cases were applied to both components separately.

The material properties shown in Table 1 were used for Ti-6Al-4V and UHMWPE in the simulation study using SolidWorks 2008.

4. Results and discussions

For the case of the articulating cylindrical head the 34.298 N forces was applied to the structure. The von Mises stress ranged from 2.524 Pa to 3.926e+005 Pa. The yield strength of Ti-6Al-4V is 8.80e+008 Pa; the von Mises stress is considerably below the yield strength as reported by the CAD software (Figure 3).

Table 1. Material properties [10-11]

Material	Properties	
Ti-6Al-4V	Elastic Modulus (Pa)	1.138e+011
	Poisson's Ratio	0.342
	Mass Density (kg/m^3)	4300
	Yield Strength (Pa)	8.80e+008
UHMWPE	Elastic Modulus (Pa)	1.2e+009
	Poisson's Ratio	0.46
	Mass Density (kg/m^3)	952
	Yield Strength (Pa)	2.1e+007

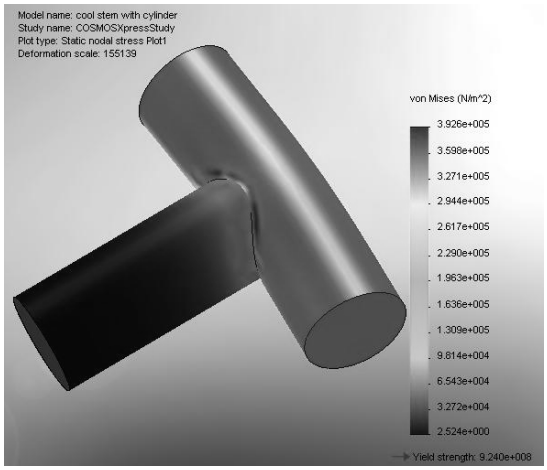


Figure 3. Articulating cylindrical head: von Mises stress

The displacement calculation yielded a maximum value of 1.157e-008 m. This is insignificant and confirms that the titanium alloy, Ti-6Al-4V is an acceptable choice for implant (Figure 4).

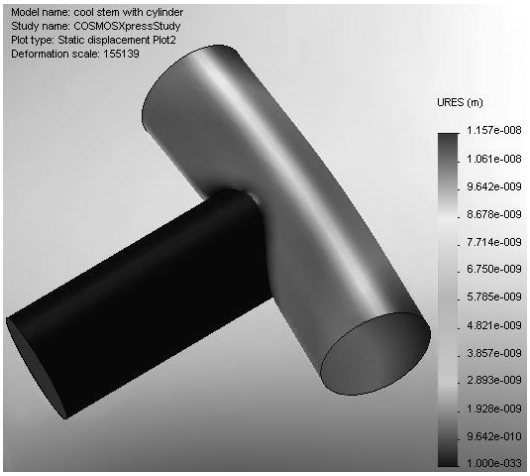


Figure 4. Displacement in articulating cylindrical head

For the case of the UHMWPE socket the Hertz P_{max} was applied to the inside structure of the socket to simulate the interaction of the articulating cylindrical head on the socket material. The tension case was analyzed since the Hoop stress predicted that this action would cause deformation in the cup area. The von Mises stress ranged from 5.995e+001 Pa to 1.309e+008 Pa. The yield strength of UHMWPE is 2.1e+007 Pa; the Von Mises stress is considerably above the yield strength as reported by the CAD software (Figure 5). This means that the UHMWPE will deform permanently in the plastic region. The displacement calculation produced a maximum value of 2.144e-003 m. This is significant and confirms that the UHMWPE will deform as shown in Figure 6. This confirms the Hoop stress calculations and requires an improvement in the articulating cup design.

Strain analysis using the CAD software was also investigated on the UHMWPE cup since it was the bottleneck in the design. The maximum strain was 8.685e-002, as shown in Figure 7, which also confirms that deflection calculations carried out using the software.

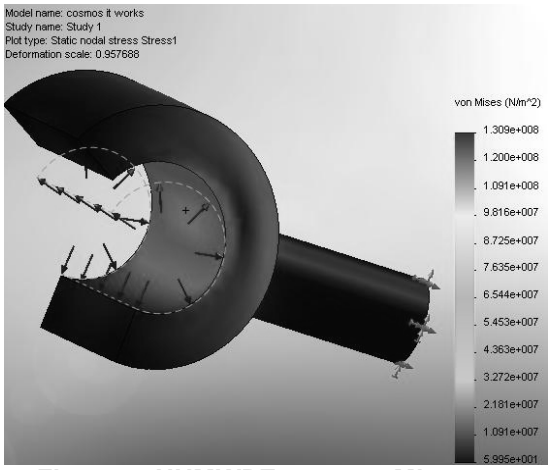


Figure 5. UHMWPE cup von Mises stress

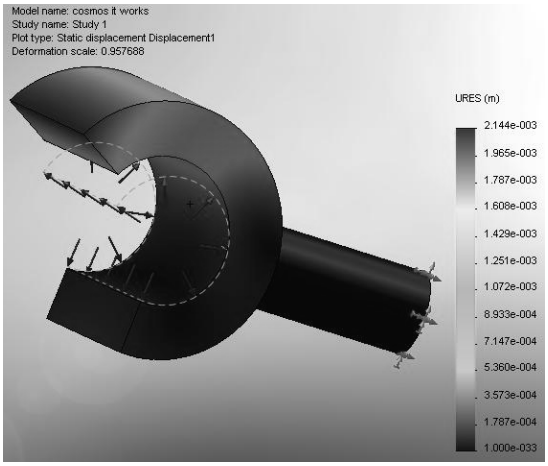


Figure 6. UHMWPE cup displacement

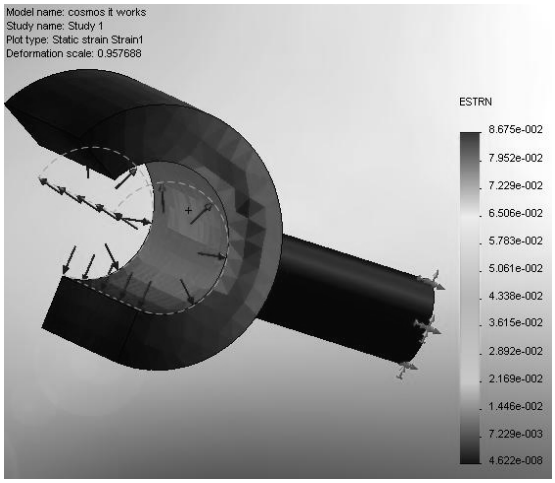


Figure 7. Strain on UHMWPE Cup

5. Conclusions

Based on the analysis of the proposed design of a PIP joint using theories such as Hertz

contact stress and Hoop stress as well as modeling the proposed joint design within SolidWorks 2008, several lessons were learned. The use of Ti-6Al-4V for the articulating cylindrical head was a good choice biomechanically because it is a biocompatible material that does not deteriorate in the body. The material will also withstand the stresses imposed on it by normal use. The Ti-6Al-4V stems will have a thin layer of beaded hydroxyapatite formula to promote bone re-growth after implantation. The UHMWPE cup was the weakest part of the design. This was observed both in the Hoop stress calculation and the CAD modeling analysis. The tension case proved to be the test that the UHMWPE needed to be examined. However, the CAD software analysis assumed that no other supporting structures (tendons, ligaments, etc.) are providing stability to the joint. In reality, 80% of supporting structures provided by the ligaments and tendons around the joint are providing additional support to the joint after surgery. Perhaps this could be enough to prevent dislocation of the UHMWPE cup.

However, the design does provide the full 100° of flexion and extension range of motion as does the natural PIP joint. The implant design does consider dislocation prevention by having an articulating ball and socket like design, though further investigation is needed for the socket design. Overall, with minor adjustments to the socket design, the proposed PIP joint design could prove successful in future treatment of rheumatoid arthritis [5].

The study on alternate cup materials proved useful as well. The Ti-6Al-4V and Co-Cr were found to be very capable of replacing the UHMWPE as the socket material. However, the wear rate of the Ti-6Al-4V articulating head on either Ti-6Al-4V or Co-Cr socket could prove to be too high for implant use. Most other joint replacements utilize an UHMWPE meniscus to cushion the joint during normal use and to eliminate metal on metal contact to reduce material wear. It is for these reasons that the UHMWPE should still be considered the top choice for the socket material. The alternative material study also proved that silicone is not suitable for the socket material since it deforms easily under load.

5.1. Future research

The PIP implant design will be machined by local professional machine shop using titanium alloy and UHMWPE. Three actual implants will be implanted into cadaveric digits (Mercer Medical School Gross Anatomy Lab) for range of passive motion tests using a Rapid Upper Limb Assessment (RULA) tools and standardizing the implant procedures. Besides that, three actual PIP implants will be tested using MTS at Mercer University Prosthetic Lab for determining the wear of the UHMWPE meniscus.

Eventually, the PIP implant will be ready for actual human trial tests in third world countries. The data from the human trials will be submitted to FDA for approval in the United States.

6. References

[1] H. E. Ash and A. Unsworth, “Design of a Surface Replacement Prosthesis for the Proximal Interphalangeal Joint”, The Journal of Engineering and Medicine, Volume 214.2 (2000), pp.151-163.

[2] Hand Anatomy|Eorthopod.com, "Eorthopod.com - by the Medical Multimedia Group, L.L.C. Web. 01 Aug.2010: <http://www.eorthopod.com/content/hand-anatomy>.

[3] N. K. Fowler and A. C. Nicol, “Measurement of External Three-Dimensional Interphalangeal Loads Applied During Activities of Daily Living”, Clinical Biomechanics, Volume 14 (1999), pp. 646-652.

[4] C. Sollerman, “How I do it – MP Joint Arthroplasty”, International Congress Series, Volume 1295 (2006), pp. 144-153.

[5] P. E. Blazar, “Metacarpophalangeal Arthroplasty in Rheumatoid Arthritis”, The Orthopedic Journal at Harvard Medical School (2001): http://www.orthojournalHMS.org/volume5/manuscripts/ms14.htm. 8 December 2009.

[6] T. P. Schweitzer, “The Terminal Tendon of the Digital Extensor Mechanism: Part I, Anatomic Study”, Journal of Hand Surgery, Volume 29, Issue 5, September 2004.

[7] Shigley, J. E., R. M. Charles, and G. B. Richard, "Chapter 4 Load and Stress Analysis", Mechanical Engineering Design, McGraw-Hill, New York, 2004.

[8] P. Pallav, "Contact Stress", Dental Materials, Home. Web. 1 Aug. 2010: <http://www.dentalmaterials.nl/wear/basics/hertz/hertz.htm>.

[9] Shigley, Joseph Edward., Charles R. Mischke, and Richard G. Budynas. "Chapter 4 Load and Stress Analysis", Mechanical Engineering Design, New McGraw-Hill, York, 2004, pp. 149-51.

[10] "ASM Material Data Sheet", ASM. Web. 01 Aug. 2010:<http://asm.matweb.com/search/SpecificMater ial.asp?bassnum=MTP641>.

[11] Kurtz, S. M., "Chapter 1: A Primer on UHMWPE", UHMWPE Biomaterials Handbook Ultra High Molecular Weight Polyethylene in Total Joint Replacement and Medical Devices, Elsevier/Academic, Amsterdam, 2009.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

A Calculus for Modeling Security and Mobility in Multi-Agent Systems

Jean Gourd
Louisiana Tech University
jgourd@latech.edu

Dia Ali
University of Southern Mississippi
Dia.ali@usm.edu

Abstract

Although the concept of mobile computation is not new, its implementation invariably introduces a variety of issues that must be addressed. Mobile agents, autonomous entities that typify mobile code, typically cross the administrative boundary of the systems they inhabit thereby raising numerous security concerns. It often behooves designers of such *mobile* systems to first model them in order to ascertain a variety of important characteristics such as validity, scalability and security.

In this paper, we present API-S, a calculus for modeling security and mobility in multi-agent systems. Fundamentally, API-S extends the pi-Calculus; however, it directly extends the API Calculus by integrating numerous constructs and primitives that combine to provide a powerful way of modeling a variety of security methods, protocols and techniques (including secure communication) within groups of interacting intelligent mobile agents. We subsequently illustrate the calculus through a variety of simple examples and models in order to show how it can ultimately provide a formal method that is capable of accurately modeling realistic distributed computational systems.

1. Introduction

Over the past decade, the Internet has grown tremendously. About one-sixth of the world's population [2, 12] regularly uses the Internet to perform everyday tasks. Internet-ready devices such as cellular phones and PDAs (personal digital assistants) are becoming ubiquitous. The traditional method of connecting to a network via cables is being largely replaced with wireless methods of connectivity, thus allowing users an untethered connection to cyberspace. Mobile computing is indeed becoming more prevalent due to the increasing need to maintain connectivity while "on the go." Considering this progressive direction toward a "mobile society," maintaining security in such an environment is patently crucial.

According to the National Institute of Standards and Technology, the cost of software bugs could reach $60 billion a year [19]. Providing performance guarantees and verifying system correctness greatly enhances the reliability of these systems [19]. In the computing sciences, modeling real-world systems prior to formally actualizing them provides essential insights that allow near-unlimited experimentation and "tinkering." Valuable assessments, such as the system's scalability and performance, can lead to better and more robust systems that are ultimately derived from the model. At the most basic level, a model of the system can be used to prove its correctness. Modeling tools provide a useful way of reasoning about complex systems and assist in answering very important questions such as:

(1) Does the system perform as intended?
(2) Is the system's performance acceptable?
(3) How will the system react to changes?
(4) How does the system comparatively perform?

Modeling systems that exhibit aspects of mobility introduces a level of complexity due to the inherent dynamic topology of the environment [11]. Typically the models are complex because so many variables are involved. Often, modeling such systems is intractable and cannot be entirely solved using computers; therefore, they tend to be modeled at higher levels of abstraction [7].

1.1. The Intelligent Agent

Intelligent agents are widely used in distributed systems, particularly those that employ the use of mobile code. At its core, the intelligent agent is an autonomous entity that performs tasks on behalf of users. Wooldridge and Jennings further characterize an intelligent agent by imposing such qualities as social ability, reactivity and pro-activeness [20]. As such, an agent can communicate with other agents, alter its behavior in response to environmental changes and induce changes by initiating behavioral patterns.

The intelligence an agent possesses is a computational intelligence that attempts to impart an agent with characteristics similar to humans; for example, an agent may possess motivation and intention when performing tasks. Most often, such an intelligence is implemented by providing the agent with a set of rules to follow while simultaneously providing an ability to learn. Numerous techniques within the field of artificial intelligence such as fuzzy logic and artificial neural networks are typically employed to accomplish this.

1.2. Mobile Agents

Mobile agents are a subclass of intelligent agents that possess an ability to wholly migrate from one host to another. This inherently introduces a variety of security concerns, most importantly that mobility constitutes a crossing of the administrative boundary of systems. Furthermore, the mobile agent resumes execution on remote systems once it has migrated which may pose risks to the remote environment or to the agent; for example, the agent's execution must be restricted, its potentially proprietary algorithms and data must be protected and so on.

Generally speaking, security threats become increasingly significant in the presence of mobile code. Mobile code violates all typical security assumptions to the point that even single-user systems require hefty security. Mobile agents are particularly vulnerable when situated on a remote host. Typically, the mobile agent must give access to its code, potentially its state, and at times its data, which renders it vulnerable to exploitation.

1.3. Multi-Agent Systems (MAS)

Most agent-based systems exhibit an intrinsic collaboration among agents. In many cases, each agent cannot solve the problem unaided, and the system relies on multiple agents. There typically is no global control mechanism as agents work on their own autonomously. Often data are decentralized and distributed over a wide network and computation is asynchronous; multiple agents carry out tasks simultaneously. MAS quickly become complicated and succinctly modeling them can indeed prove to be a challenge. Nonetheless MAS are interesting because the agents that comprise them often exhibit natural groupings and an emerging order. In many cases MAS have been introduced to allow for broader scope and increased scalability (see works in [8, 9, 13, 15, 16, 17, 18]).

In this paper, we introduce the API-S Calculus, a formal modeling tool that can be used to model the complexity of agents within a MAS, while providing powerful mechanisms to accurately model the mobility and natural grouping behavior of agents in addition to the security implemented within the system. In Section 2 we introduce the API-S Calculus. In Section 3 we illustrate the calculus via several

simple examples, and in Section 4 we conclude and provide future directions.

2. The API-S Calculus

The API-S Calculus (API-S) extends Rahimi's API Calculus [14] by integrating primitives and constructs necessary to capably model the security of agents within a MAS. API-S forms a foundation on which additional constructs can be built, thus providing the necessary tools to extend the breadth of the types of systems the calculus can model.

Two constructs (Ω-term and Ω-process) are introduced that provide useful primitives necessary to model a wide variety of cryptographic protocols. The *milieu*, a semantic abstraction that provides a powerful way to model families of agents, is inherited from the API Calculus, and its definition is only slightly expanded upon in order to more capably model the security of the remote agent host. In this way, we can model a group of agents working cooperatively on a similar task or even a group of agents residing on the same host involved in secure communication. Furthermore, API-S provides a method to model secure process communication across groups of agents.

2.1. Syntax

Terms are defined as follows:

R,T	::=	$x,y,z,...$	name
	|	$a,b,c,...$	fact or rule
	|	$f(x,y,z,...)$	function
	|	R^{Ω},T^{Ω}	Ω-term

A name may be a channel or a variable name. A fact or a rule can be added to a knowledge base, sent to a different process or received from a different process. A function may have l parameters.

Variables u,v,w are used to range over names, facts and rules. The abbreviations $\boldsymbol{u}$ and $\boldsymbol{T}$ are used in place of the tuples $u_1,u_2,...,u_l$ and $T_1,T_2,...,T_l$ respectively.

An Ω-term compartmentalizes cryptographic primitives for use in modeling an array of cryptographic protocols and techniques.

Processes are defined as follows:

P	::=	0	no action
	|	$\alpha.P$	action prefix
	|	P_1+P_2	summation prefix
	|	$[T=R]P_1:P_2$	conditional prefix
	|	υxP	name restriction
	|	$(K_i)P$	knowledge unit restriction
	|	$!P$	replication
	|	$D<\boldsymbol{T}>$	constant
	|	$M<<P>>N$	listener
	|	P^{Ω}	Ω-process

Processes are denoted by $P_1,P_2,...,P_n$ and $Q_1,Q_2,...,Q_n$. The no action process is denoted by 0 and performs internal computation. The prefix α is called the action prefix. The expression $\alpha.P$ performs the action α and then behaves as P. The summation process P_1+P_2 behaves as either P_1 or P_2. Terms are identified by T and R. The process $[T=R]P_1:P_2$ is a conditional process such that the equality of T and R is not a strict syntactic identity. When $P_2=0$, we abbreviate the conditional process to $[T=R]P_1$. The expression υxP makes a new private name x local to P then behaves as P. The knowledge unit restriction $(K_i)P$ makes a new private knowledge unit K_i local to P. The replication process $!P$ implies $(P|P|...|P)$; for example, $P_1|P_2$ consists of two processes P_1 and P_2 acting in parallel and perhaps independently. Process synchronization (i.e., of an output action of P_1 at some output port $\boldsymbol{x}$ with an input action

of P_2 at some input port x) is possible resulting in a silent (τ) action.

The constant D, where $\boldsymbol{L}$ indicates a tuple of processes or terms, is a function whose parameters can be processes or other functions. For example, consider $\boldsymbol{x}L$ to be an output prefix that sends a term or process L on channel x and $x(L)$ to be an input prefix that receives a term or a process L over channel x. Now consider the expression $\boldsymbol{x}P.Q|x(X).X$. Once the interaction between the two processes has occurred, the resulting process is then $Q|P$. This may be interpreted as a process $x(X).X$ waiting for X to be sent along channel x, thus defining its subsequent behavior. In this case, the process P is sent along channel x in the expression $\boldsymbol{x}P$; the process $x(X).X$ then simply behaves as P once the interaction has taken place.

In $M<<P>>N$, process P is associated with (and listening to) milieus M and N. This allows a process to be part of more than one communication group. Process P may receive messages broadcast by processes within milieus M and N and may broadcast messages to any processes within M and N.

An Ω-term U includes primitives for cryptographic support. Ω-terms are defined as follows:

R^{Ω}, T^{Ω}	::=	0	zero
	|	(M,N)	pair
	|	$suc(M)$	successor
	|	$H(M)$	hashing
	|	$\{M\}_K$	shared key encryption
	|	$\{[M]\}_K$	public key encryption
	|	$[\{M\}]_K$	private key signature
	|	K^+	public key part
	|	K^-	private key part

Constructs for pairing and numbers, (M,N), 0 and $suc(M)$, are added for convenience in modeling numerous cryptographic protocols. The successor term $suc(M)$ implies $M+1$. For example, $suc(0)=1$ and $suc(suc(0))=2$.

$H(M)$ represents the hashing of some message M (which can be a key). Furthermore, we assume that H is a one-way function (i.e., it cannot be inverted) and that it is free of collisions. We write $\{M\}_K$ to represent the encryption of a message M with a key K using a shared key cryptographic system (e.g., DES [4]). Public key encryption (e.g., RSA [10]) is modeled as $\{[M]\}_K$, which illustrates some message M encrypted under some key K. Often, we wish to encrypt using some public key and typically write $\{[M]\}_{K+}$ to denote this behavior. Accordingly, if K represents a key pair, then K^+ corresponds to its public part and K^- corresponds to its private part. We write $[\{M\}]_K$ to denote a message M that is digitally signed with key K.

To support the formal modeling of conditional processes regarding authoritative cryptographic protocols, the Ω-process is introduced. The digital signature of terms, processes, and agents is capably modeled in API-S, in addition to common cryptographic schemes. Ω-processes are defined as follows:

P^{Ω}	::=	$let\ (x,y)=M\ in\ P$	pair splitting
	|	$case\ M\ of\ 0:P, suc(x):Q$	integer case
	|	$case\ L\ of\ \{x\}_N\ in\ P$	shared key decryption
	|	$case\ L\ of\ \{[x]\}_N\ in\ P$	public key decryption
	|	$case\ L\ of\ [\{x\}]_N\ in\ P$	signature check

In $let\ (x,y)=M\ in\ P$, the variables x and y are bound in P. It behaves as $P\{NL/xy\}$ if M is the pair (N,L). In $case\ M\ of\ 0:P, suc(x):Q$, the variable x is bound in Q. It behaves as P if M is 0 or as $Q\{N/x\}$ if M is $suc(N)$. The variable x is bound in P in all three of the following processes:

- *case L of* $\{x\}_N$ *in P* attempts to decrypt *L* with the key *N*. If *L* is a cypher text of the form $\{M\}_N$, then the process behaves as $P\{M/x\}$.
- *case L of* $\{[x]\}_N$ *in P* attempts to decrypt *L* with the key *N*. If *L* is a cypher text of the form $\{[M]\}_N$, then the process behaves as $P\{M/x\}$. If *N* is a private key K^-, then *x* is bound to *M* such that $\{[M]\}_{K+}$ is *L*, if such an *M* exists.
- *case L of* $[\{x\}]_N$ *in P* attempts to decrypt *L* with the key *N*. If *L* is a cypher text of the form $[\{M\}]_N$, then the process behaves as $P\{M/x\}$. If *N* is a public key K^+, then *x* is bound to *M* such that $[\{M\}]_{K-}$ is *L*, if such an *M* exists.

We should note that several assumptions about cryptography are made as pointed out in [1]: (1) the only way to decrypt an encrypted message is to know the appropriate key; (2) an encrypted message does not reveal the key that was used to encrypt it; and (3) the decryption algorithm used can always detect whether a message was encrypted with the expected key (i.e., there is adequate redundancy in the message).

2.2. Knowledge Units

A knowledge unit $K_1,K_2,...,K_n$ consists of a knowledge base (that is composed of rules) and a set of facts. A knowledge unit reacts to new facts added to its fact list. K^i denotes the set of knowledge units belonging to process P_i. Knowledge units provide a level of intelligence for agents and are defined as follows:

K	::=	0	empty unit
	|	r	rule
	|	K_1+K_2	knowledge unit summation

The empty knowledge unit, one with no rules or facts, is denoted by *0*. A knowledge unit may consist of a single rule. The summation K_1+K_2 indicates that both knowledge units K_1 and K_2 react to a fact simultaneously, essentially behaving as a single knowledge unit.

2.3. Milieus

A milieu is a bounded place (also called an environment) in which processes reside and computations take place. Although milieus are similar to ambients in the ambient calculus (see [3] and [5]), there is a great difference between the two: milieus are not a basic unit of a system; rather, they represent an environment in which processes can join together to form a new computational unit [14]. The existence of separate locations is represented by a topology of these boundaries.

A milieu is surrounded by a border that must be traversed in order to join or leave it. We will show that communication can indeed occur through the milieu boundary via an unnamed environment channel. An entire milieu can move, taking with it its entire contents (i.e., all the processes and other milieus within it). Milieus are well suited to address the characteristics of the natural grouping and security of the system.

M	::=	0	empty milieu
	|	$\beta.M$	action prefix
	|	$M[O]$	ownership
	|	$M[O_1\|O_2]$	parallel
	|	M_1+M_2	summation of milieus

An empty milieu is denoted as *0*. The variables $O_1,O_2,...,O_n$ are used to range over processes and milieus. $M[O]$ is a milieu in which process or milieu *O* exists. A milieu may consist of other milieus or processes behaving in parallel; for example, $M[O_1|O_2]$. The expression M_1+M_2 indicates that milieu *M* is generated by the merging of milieus M_1 and M_2. The prefix β

is an action prefix. The expression $\beta.M$ performs the action β and then behaves as M.

$M[O]$ exhibits a tree structure induced by processes and the nesting of milieu brackets (e.g., $M[P_1|...|P_p|M_1[...]|...|M_q[...]]$). In API-S process mobility is represented as the crossing of milieu boundaries; however, interaction between processes can cross the milieu boundary.

2.4. Actions

As in the API Calculus, terms and processes may be present within actions. Knowledge unit actions include the receiving and sending of knowledge units and the adding and dropping of facts and rules. Milieu actions include the ability to join or leave a milieu, and the ability to open a milieu boundary. The process action prefix is denoted by α; the milieu action prefix is denoted by β.

Let A be the set of all α-actions in the calculus:

- τ is a silent, internal action.
- $x(\boldsymbol{L})$ is an input prefix where x is the input port or channel of a process which contains it and $\boldsymbol{L}$ is a tuple of processes or terms. The process $x(\boldsymbol{L}).P$ inputs an arbitrary number of terms or processes $\boldsymbol{L}$ on channel x and then behaves as $P\{\boldsymbol{L}'/\boldsymbol{L}\}$. All free occurrences of the names $\boldsymbol{L}$ in P are bound by the input action prefix $x(\boldsymbol{L})$ in P.
- $\boldsymbol{xL}$ is an output prefix where x is the output port or channel of a process which contains it and $\boldsymbol{L}$ is a tuple of processes or terms. The process $\boldsymbol{xL}.P$ outputs an arbitrary number of terms or processes $\boldsymbol{L}$ on channel x and then behaves as P.
- $(\boldsymbol{K})P$ makes the tuple of knowledge unit names $\boldsymbol{K}$ local to P.
- $K_i\langle\boldsymbol{a}\rangle(\boldsymbol{R})$ is a knowledge unit call. The expression $K_i\langle\boldsymbol{a}\rangle(\boldsymbol{R}).P$ calls the knowledge unit K_i, passing a list of facts $\boldsymbol{a}$. The result of this call is placed in $\boldsymbol{R}$. All free occurrences of $\boldsymbol{R}$ in P are bound by the prefix $K_i\langle\boldsymbol{a}\rangle(\boldsymbol{R})$ in P.
- $x(\boldsymbol{K})$ is an input prefix where $\boldsymbol{K}$ is a tuple of knowledge units that is received from a process via channel x. The expression $x(K_1).P$ receives a knowledge unit K_1 on channel x and then behaves as P.
- $\boldsymbol{xK}$ is an output prefix where $\boldsymbol{K}$ is a tuple of knowledge units that is sent by a process via channel x. The expression $\boldsymbol{x}K_1.P$ sends a knowledge unit K_1 on channel x and then behaves as P.
- $K_i(\boldsymbol{a})$ adds the tuple $\boldsymbol{a}$ to the fact list of K_i (if $\boldsymbol{a}$ is a tuple of facts) or to the rule base of K_i (if $\boldsymbol{a}$ is a tuple of rules). The expression $K_1(a).P$ adds a to the fact list or rule base of K_1 depending on the type of a.
- $K_i\boldsymbol{a}$ drops the tuple $\boldsymbol{a}$ from the fact list of K_i (if $\boldsymbol{a}$ is a tuple of facts) or from the rule base of K_i (if $\boldsymbol{a}$ is a tuple of rules). The expression $K_ia.P$ drops a from the fact list or rule base of K_i depending on the type of a.
- *join* $m.P$ allows process P to join milieu m and then behave as P inside of m.
- *leave* $m.P$ allows process P to leave milieu m and then behave as P outside of m.
- $\langle\boldsymbol{L}\rangle$ is a broadcast output prefix such that the tuple of processes or terms $\boldsymbol{L}$ is broadcast to the surrounding milieu. Any processes within the milieu in which $\boldsymbol{L}$ was broadcast may receive it. There is no notion of a channel; however, one may think of the milieu as an environment channel, and processes defined to be listeners of, or associated with, a milieu can listen or send messages on this channel. If there is more than one process listening, one is chosen in a non-deterministic manner; naturally, the diffusion can be implemented in such a way that if there are several processes listening, all of them receive the message. This action provides support for general broadcasts.
- $(\boldsymbol{L})$ is a broadcast input prefix such that the tuple of processes or terms $\boldsymbol{L}$ is received upon the condition that P is within the milieu that initially broadcast $\boldsymbol{L}$ or that P is

listening to the milieu. All free occurrences of the names ***L*** in *P* are bound by the input listening prefix ***(L)*** in *P*.

Let *B* be the set of all *β*-actions in the calculus:

- *join m.M* indicates that milieu *M* joins milieu *m* and then behaves as *M* inside of *m*.
- *leave m.M* indicates that milieu *M* leaves milieu *m* and then behaves as *M* outside of *m*.
- *open.M* indicates that the boundary surrounding milieu *M* is dissolved; *M* ceases to exist. Any processes or other milieus that were inside of *M* behave as if they do not belong to *M*.

β-actions assist in modeling groups of cooperating agents within larger groups. For example, a large group of agents may naturally divide into several cliques, and we may wish to model this behavior.

2.5. Other Features

API-S provides a plethora of other necessary and useful features and constructs that assist in modeling a system in the calculus (e.g., binding of variables, substitution and convertibility of expressions and elements, broadcasting, triggers, useful abbreviations, structural congruence, and reduction). What follows is a brief summary of these features through illustrative examples. For a thorough introduction and detailed discussion of these features, the reader is referred to [6].

3. Simple Examples of API-S

In this section, we illustrate the use of API-S as a modeling tool through a variety of simple examples.

3.1. Client-Server-Printer Example

Figure 1. The client-server-printer example

Server *S* shares a link *x* with printer *P*; client *C* shares a link *y* with the server. *C* wishes to send a document *d* to *P* for printing. In this example, we illustrate the passing of the communication link *x* from *S* to *C* so that the client can use this link to send its document. The act of printing the document is modeled by the function *F()*, an internal function in *P*. We define the system illustrated in Figure 1 as follows:

$S = \boldsymbol{y}x.S'$

$P = x(d').F(d').P'$

$C = y(x').\boldsymbol{x'}d.C'$

The behavior is then:

$\boldsymbol{y}x.S'|x(d').F(d').P'|y(x').\boldsymbol{x'}d.C'$

$\rightarrow S'|x(d').F(d').P'|(\boldsymbol{x'}d.C')\{x/x'\}$

$\rightarrow S'|(F(d').P')\{d/d'\}|C'$

$\rightarrow S'|P'|C'$

3.2. Passing a Knowledge Unit

Consider two agents A_1 and A_2 that share a link *x*. Agent A_1 possesses a knowledge unit K_1 and agent A_2 possesses a knowledge unit K_2. Agent A_1 wishes to transfer its knowledge unit K_1 to agent A_2 using a private key *S* they both share. This example (illustrated in Figure 2) can be simply modeled in the calculus as follows:

$A_1 = \boldsymbol{x}\{K_1\}_S.A'_1$

$A_2 = x(K).case\ K\ of\ \{y\}_S\ in\ A'_2$

The behavior is then:

$\boldsymbol{x}\{K_1\}_S.A'_1|x(K).case\ K\ of\ \{y\}_S\ in\ A'_2$

$\rightarrow A'_1|(case\ K\ of\ \{y\}_S\ in\ A'_2)\{K_1/K\}$

$\rightarrow A'_1|A'_2\{K_1/y\}$

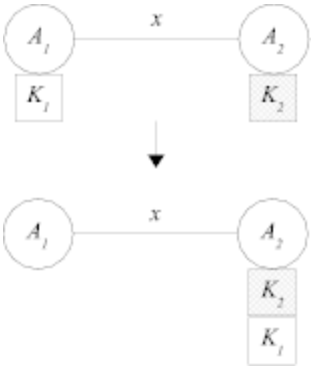


Figure 2. Passing a knowledge unit

The result is such that agent A_2 now possesses knowledge units K_1 and K_2.

3.3. Migrating Agent with Data Request

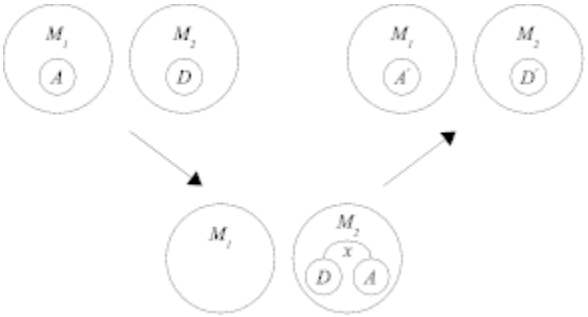


Figure 3. Agent migration with data request

We wish to illustrate that API-S can capably model a real world example, albeit a fairly simple one. In the scenario illustrated in Figure 3, there exists two remote hosts M_1 and M_2 (modeled as milieus), one agent A, and one database D (a process). A, initially at host M_1, will migrate to M_2, query D with a request for information q, receive a response r from D, and return home. A simple way to implement this scenario in the calculus is as follows:

$M_1 = [A]$

$M_2 = [D]$

$A = leave\ M_1.join\ M_2.\mathbf{x}q.x(r').leave\ M_2.join\ M_1.A'$

$D = x(q').\mathbf{x}r.D'$

The behavior of the system becomes:

$M_1[leave\ M_1.join\ M_2.\mathbf{x}q.x(r').leave\ M_2.join\ M_1.A'] | M_2[x(q').\mathbf{x}r.D']$

$\rightarrow M_1[] | M_2[\mathbf{x}q.x(r').leave\ M_2.join\ M_1.A' | x(q').\mathbf{x}r.D']$

$\rightarrow M_1[] | M_2[x(r').leave\ M_2.join\ M_1.A' | (\mathbf{x}r.D')\{q/q'\}]$

$\rightarrow M_1[] | M_2[(x(r')\{r/r'\}.leave\ M_2.join\ M_1.A' | D']$

$\rightarrow M_1[A'] | M_2[D']$

For more examples, the reader is referred to [6].

4. Conclusions and Future Directions

In this paper, we introduced API-S, a calculus for modeling security and mobility in MAS. API-S offers powerful constructs that provide support for modeling cryptographic protocols and various security techniques often unique to the mobile agent paradigm and MAS. We have illustrated the calculus through a variety of simple examples and models in order to demonstrate how it can ultimately provide a formal modeling tool that is capable of accurately modeling realistic distributed computational systems.

There remain several areas for future work that would prove useful in improving the calculus: (1) extending the calculus by integrating new constructs that provide support for modeling new mobile agent security techniques as they are proposed; and (2) the design and implementation of an automated tool to computationally verify expression syntax, reduce expressions and evaluate modeled

systems. This would essentially be an analytical modeling framework for API-S coded models.

References

[1] M. Abadi and A. D. Gordon. A calculus for cryptographic protocols: The spi calculus. In Fourth ACM Conference on Computer and Communications Security, pages 36–47. ACM Press, 1997.

[2] US Census Bureau. Us and world population clocks. Internet Web site, November 2006. http://www.census.gov.

[3] L. Cardelli, G. Ghelli, and A. D. Gordon. Types for the ambient calculus. *Information Computation*, 177(2):160–194, 2002.

[4] D. Coppersmith. The data encryption standard (des) and its strength against attacks. *IBM Journal of Research and Development*, 38(3):243–250, 1994.

[5] A. D. Gordon. Part iii: The ambient calculus. Presented at the International Summer School on Foundations of Security Analysis and Design (FOSAD 2000) in Bertinoro, September 2000.

[6] J. Gourd. API-S calculus: formal modeling for secure mobile intelligent agent systems. PhD thesis, The University of Southern Mississippi, Hattiesburg, MS, USA, August 2007.

[7] L. O. Hertzberger. The importance of modeling for computational science. In J. Gerbrandy, M. Marx, M. de Rijke, and Y. Venema, editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*. Amsterdam University Press, 1999.

[8] N. R. Jennings, K. Sycara, and M. Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38, 1998.

[9] S. M. Koriem. Development, analysis and evaluation of performance models for mobile multi-agent networks. *Computing Journal*, 49(6):685–709, 2006.

[10] RSA Laboratories. Pkcs#1 v2.1: Rsa cryptography, standard draft 2, January 2001.

[11] E. F. A. Lima, P. D. L. Machado, F. R. Sampaio, and J. C. A. Figueiredo. An approach to modelling and applying mobile agent design patterns. *SIGSOFT Software Engineering Notes*, 29(3):1–8, 2004.

[12] J. Nielsen. One billion internet users. Internet Web site, December 2005.

[13] A. Pridgen and C. Julien. A secure modular mobile agent system. In *SELMAS '06: Proceedings of the 2006 international workshop on Software engineering for large-scale multi-agent systems*, pages 67–74, New York, NY, USA, 2006. ACM Press.

[14] S. Rahimi. API-Calculus for Intelligent-Agent Formal Modeling and its Application in Distributed Geospatial Data Conflation. PhD thesis, The University of Southern Mississippi, Hattiesburg, MS, USA, May 2002.

[15] S. Rahimi and N. F. Carver. A multi-agent architecture for distributed domain-specific information integration. In *HICSS '05: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4*, page 113.2, Washington, DC, USA, 2005. IEEE Computer Society.

[16] T. Rorie, J. Burge, and A. C. Esterline. Formal modeling of multi-agent systems: An application of the pi-calculus. In M. Jamshidi, D. Kaufman, and N. Vadiee, editors, *Proceedings of ACE-PURSUE STUDENT CONFERENCE, Advances in Research and Education in Science, Mathematics and Engineering*, April 1999. Albuquerque, NM.

[17] T. Rorie and A. C. Esterline. Formal modeling of multi-agent systems using the pi -calculus and epistemic logic. In *NASA-URC98*, Huntsville, AL, 1998.

[18] R. Tatikunta, S. Rahimi, P. Shrestha, and J. Bjursel. Tragent: A multi-agent system for stock exchange. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)*, pages 505–509, 2006.

[19] J. Woodcock. First steps in the verified software grand challenge. *Computer*, 39(10):57–64, 2006.

[20] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Assessing Operator Workload for a Fluid Powered Rescue Crawler

Antonio Lee and Steven Jiang
North Carolina Agricultural & Technical State University
xjiang@ncat.edu

Abstract

Workload has proven to be an important factor in the attempts to investigate human performance and its contributing factors. This assessment technique has the potential to further assist in the understanding of how certain tasks will affect operators performing tasks whether physical or cognitive. The goal of this research is to assess the human operator workload when different control types are used for the rescue crawler. Results indicated that there was a significant difference in the amount of effort required from the human operator using traditional control types versus the haptic control types, which suggests the traditional design is a good application to utilize when operating a rescue crawler.

1. Introduction

Recently robots have begun to be utilized in many critical application domains that are considered to be too dangerous for humans such as search and rescue missions. The primary concern deals with measuring how well the human(s) and robot(s) perform as a team or the human robotic interaction (HRI) [1]. There are many well- known task measures used to investigate system performance ranging from subjective ratings used for investigating the quality of effort involving all stakeholders both direct and indirect to the appropriate utilization of mixed measures [2]. In this research quantitative performance measures are utilized to assess the effectiveness and efficiency of the operator performing a task. The efficiency of the rescue crawler was considered in the form of amount of time to complete the task.

Historical research has demonstrated that operator performance and workload are significantly affected by various controls that are required for task performance [3, 4]. The term workload may refer to mental or cognitive workload as well as physical and temporal workload. In general, high workload "can lead to reduction of vigilance as the person struggles to maintain accuracy and judgment under information and time pressure" [5]. Physical workload measures play an important role in allocating tasks and should be employed not only for automatic task allocation but also in the initial HRI design and development to guarantee that the operator is not routinely over or under loaded [6]. Mental workload is strongly influenced by demands made on both long term and short term memory. In this research, the approach considered for this HRI design was the allocation of tasks

between the operator and the robotic systems in the form of physical workload.

2. Methodology

2.1 Data Collection

Data collection was completed via examining video clips of rescue robot operators interacting with the robot using both traditional controls and haptic controls. The operators were categorized as experts or novices depending on their expertise and experience with the crawler. Therefore, there were four experimental conditions in the research: expert operating a traditional control, expert operating a haptic control, novice operating a traditional control, and novice operating a haptic control. Time studies were conducted for each condition and recorded for further analysis.

2.2 Task Analysis

A task analysis was developed based on each individual step performed by the operators to document the average completion time for each movement.

2.3 Model Development

Results of the task analysis were then used to build four simulation models in the form of task networks using the Micro Saint Software [7,8,9]. Two factors studied are Expertise (expert, novice) and Control type (haptic, traditional). The tasks in each model were then categorized in the form of human and system tasks to provide compatible workload percentages from each process level. The design for the model was derived from the task analysis and ordered from top-level goals to lower level activities. After listing the top-level goals in the order of operation, the sub goals branched off from each top-level goal with its corresponding sub goals. The individual tasks for implementing sub goals were then identified. Information such as the probability of each action being taken, the amount of time it takes to perform a task, amount of expertise required to perform a task was documented.

Table 1. Decision Probabilities

	Move object		Shift around object	
	Freq.	Prob.	Freq.	Prob.
Expert	10	.67	5	.33
Novice	6	.4	9	.6

To determine the probability of an action that the operator would take to either move the object in front of it or to shift around the object, the following procedure was used. For each video clip, the level of expertise was documented as well as the decision the operator made. Table 1 shows the frequency and probability calculated for each situation. Using the information, the task network diagram was then built. Activities in these diagrams are represented by nodes and the arrows between the nodes represent the sequence of the activities.

2.4 Data Analysis

In this research, workload was measured as a ratio between the amount of effort required by human operator and the system. It was done through coding each node of the model in the form of human and system efforts (using completion time) needed to complete a certain task.

Figure 1 describes the formulation of the workload calculation proposed by Hughes [10]. It can be seen that operator workload is computed as a measure of the total cognitive and physical effort exerted by the human operator task time divided by the task time to reach the outcome of system or system goal.

Human Component

$$\sum O_T + \sum M_T = E_{HO}$$

Measure of Workload

$$E_{HO} \ / \ E_S$$

System Component

$$\sum S_T = E_S$$

Figure 1. Model formation for operator workload

Effort of the human operator (E_{HO}) was measured by the summation of the operator's physical movements (M_T) and the operator's cognitive task time (O_T) preceding the system task goal; whereas, effort of the system (E_S) was measured by the execution of the system's task time. The comparison ratio of effort exerted by the human versus the system results in the percentage of human operator workload (HO_{WRKLD}).

3. Results

Workload was computed for each of the four Micro Saint Models.

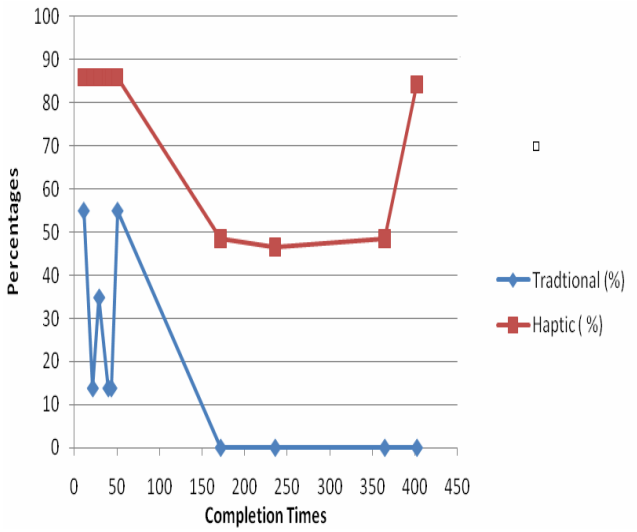


Figure 2. Novice operator (tradtional, haptic)

Figure 2 describes the workload results for novice operators using both traditional and haptic controls. It was found that approximately 30% more effort was required from the novice operators using the haptic controls than those using the traditional controls.

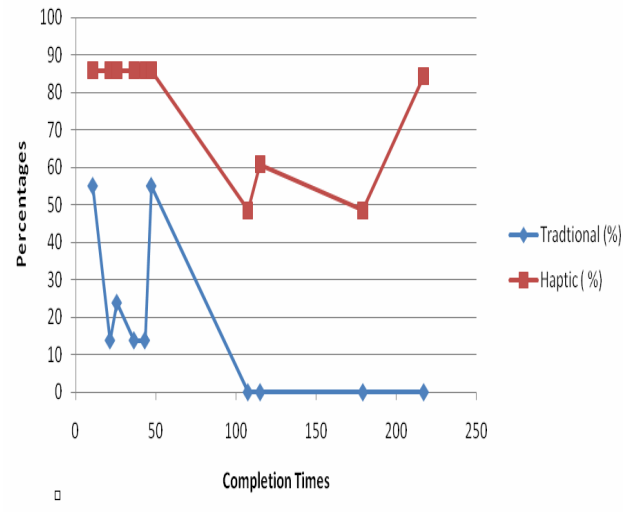


Figure 3. Expert operator (traditional, haptic)

Figure 3 describes the workload results for expert operators using both traditional and haptic controls. There is

approximately 35% more effort required from the expert operator using haptic controls than those using traditional controls through each processing level.

Table 2. Novice-Traditional control.

System Process	System Type	
	Operator	System
Set up Process	54.9	45.1
Maneuvering Process	13.8	86.2
Manipulating Process	22	78

In particular, Table 2 provides workload information about the model output for novice operator utilizing a traditional control. Results show that for maneuvering process and manipulating process, the system efforts ranges from 78% to 86% based on the completion time comparing with 13.8% to 22% for operator efforts. However, the operator workload was much higher than the system effort during the set up process, which occurred both in the beginning when turning the system on and at the end when shutting off the rescue crawler.

Table 3. Expert-Traditional control.

System Process	System Type	
	Operator	System
Set up Process	54.9	45.1
Maneuvering Process	13.8	86.2
Manipulating Process	22.9	77.1

Table 3 describes the results for the Expert operators using traditional control model. Similar pattern was found as for the novice operator using the traditional control model. In the Maneuvering and Manipulating processes the system exerts the most effort ranging from 77%-86%. In the Set up process the operator exerts more effort (54.9% vs. 45.1%) to start and shut down the system.

Table 4 reveals workload information about the novice operator using haptic control.

Table 4. Novice-Haptic control.

System Process	System Type	
	Operator	System
Set up Process	85.9	14.1
Maneuvering Process	56	44
Manipulating Process	55.1	44.9

Unlike the previous two models, in this model the operator workload was higher than the system effort for all three processes. The Maneuvering and Manipulating processes both pose a higher operator workload in the approximate range of 56%. The setup process has an initial spike in operator effort when starting up then seems to decrease in variation between the system and operator until the shut down process where the operator effort peaks again back up to 85.9%. The Set up process has the highest amount of variation between the operator (85.9%) and the system (14.1%).

Table 5. Expert- Haptic control.

System Process	System Type	
	Operator	System
Set up Process	85.9	14.1
Maneuvering Process	51.5	48.5
Manipulating Process	60.8	39.2

Table 5 provides information about the workload for the expert operator operating a rescue crawler with a haptic control model. Similar to the novice operator the traditional control model operator assumes majority of the effort when operating the rescue crawler throughout the process. The expert operator yielded a higher effort of 60.8% due to the fact that the operator was more likely to manipulate the object out of the way which required more operations for the operator to perform comparing with maneuvering around the object. Hence, even though the amount of time to complete the task was less but the amount of effort was more.

4. Discussion

An important finding of this research is that there is a higher required effort from the rescue crawler operator when utilizing the haptic controls regardless the expertise levels of the operator, which indicates that haptic control in its current form might not be an appropriate choice for the rescue crawler. Therefore, it is more appropriate to use the traditional controller due to its simplicity to learn and the small amount of effort required by the all operators using it.

5. Design Recommendations

The reason for the excessive effort required by the operator using haptic controller was due to the design issue of using two haptic devices that serve as the controls for the operator and legs of the rescue crawler. The design posed a problem because of the awkward and un-natural postures which did not agree with the natural motions of a person. Operator needs to operate the rescue crawler in the environment that requires too much effort to perform simple tasks such as walking. Therefore, the haptic control in its current form is counterintuitive and it is recommended to implement traditional based control types that agree with the operator's expectations and possibly putting haptic feedback with that specific control type to assist the operator.

6. Conclusion

Findings of this study demonstrate that the use of a simulation model can provide valuable information about the operator workload and the factors that may impact the performance. It provides an environment that mimics the real world that accounts for the unpredictable factors with the use of sample data and probabilities. In this study workload was quantified for four different models. The differences in the amount of workload were attributed to the different control types and expertise. The use of traditional controls was shown to require a significantly less amount of effort when compared to haptic control types. Therefore, traditional controls were

recommended for use in the rescue crawler to increase efficiency, work performance, and operator satisfaction.

References

[1] ANSI/AIAA. "Guide to Human Performance Measurements", AIAA, Washington, DC, 1993.

[2] T. Fong, C. Thorpe, and C. Baur, "Robot, Asker of questions." Robotics and Autonomous Systems, Carnegie Mellon University, Pittsburgh, PA, (2003), pp.1-9.

[3] D.B. Kaber, J. Riley, R. Zhou, and J.V. Draper, "Effects of visual interface design, Control interface type, and control latency on performance, telepresence, and workload in a teleoperation task." In Proc. XLVth Triennial Congress of the International Ergonomics Society, Mississippi State University Mississippi State, MS, (2000).

[4] D. Kaber, E. Onal and M. Ensley, "Design for telerobots and the effect on performance, operator situation awareness and subjective workload." Human Factors & Ergonomics in Manufacturing, Raleigh, NC, 2000, Vol. 10 (4) 409–430

[5] C. Donald, "People in Control Human Factors in Control Room Design", The Institution of Electrical Engineers, Palo Alto, *CA*, 2001, pp. 35-50.

[6] J. Adams, "Critical Considerations for Human-Robot Interface Development", Department of Computer Science, Rochester Institute of Technology, Rochester, NY 2002.

[7] D. Schunk, "Modeling with the Micro Saint Simulation Package", Winter Simulation Conference, Boulder, CO, 2000. Volume 1. pp. 274-279

[8] R. Laughery, S. Archer, B. Plott, & D. Dahn, "Connect task network modeling and the micro saint family of models.", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Diego, CA. (2000).

[9] K. R. Laughery, K.Corker, "Computer Modeling and Simulation." *Handbook of Human Factors*, New York, NY. (1997).

[10] K. Hughes, "Quantification of operator workload in existing and emergent fluid power systems", International Journal of Human Factors in Manufacturing, Department of Industrial and Systems Engineering, North Carolina Agricultural and Technical State University, Greensboro, NC, 2009.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Fuzzy Group Decision Making for Health-Care Waste Management

Mehtap Dursun, E. Ertugrul Karsak and Melis Almula Karadayi
Galatasaray University, Turkey
mdursun@gsu.edu.tr; ekarsak@gsu.edu.tr; melis.almula@ogr.gsu.edu.tr

Abstract

Evaluation of health-care waste (HCW) treatment alternatives requires the consideration of multiple subjective and conflicting criteria with the involvement of a group of experts. Constructing an efficient HCW management system, which considers environmental, economic, technical and social factors, is a highly important multi-criteria group decision making problem. This paper presents a hierarchical distance-based fuzzy multi-criteria group decision making framework for evaluating HCW treatment alternatives for Istanbul, one of the most crowded cities in Europe.

1. Introduction

The amount of waste generated in the health-care institutions is rising due to their extent of service. The waste from health-care institutions shows permanent features in the air, water and soil, and disturbs the ecological balance. Thus, disposal of medical wastes is a major concern particularly for developing countries, including Turkey. A large number of developing countries show increasing awareness regarding treatment of health-care wastes. Hence, constructing an efficient health-care waste (HCW) management system, which considers environmental, economic, technical and social factors, is of outmost importance.

In the literature, there are a few analytical studies about health-care waste management (HCWM). Mostly, health-care institutions generating the wastes are surveyed through the prepared questionnaires, field research and personnel interviews. Kilic [1] developed an integrated health-care waste management plan to minimize the risks to the health and human well-being and the environment in the Anatolian side of Istanbul. Lee et al. [2] investigated generation volume and sources, composition, and treatment and disposal methods for regulated medical wastes. Zeren [3] proposed institutional structure in order to resolve insufficient management of health-care wastes in the European side of Istanbul. Diaz et al. [4] presented some of the most common treatment and disposal methods utilized in the management of infectious health-care wastes in developing countries. Alagoz and Kocasoy [5] investigated technical information related to the available treatment technologies and compared capital investment cost, transportation/operational costs for each alternative method. Alagoz and Kocasoy [6] examined the existing situation and management practices such as the amount of health-care wastes generated, segregation procedures, collection, temporary storage and transportation of the wastes within and outside of the institution, and developed the main priorities for safe handling and transportation of health-care wastes by taking into account environmental and economic factors. Alagoz and Kocasoy [7] developed a health-care waste collection and transportation system for Istanbul. Birpinar et al. [8] carried out a survey about the current status of generation, collection, on-site handling,

storage, processing, recycling, transportation, and safe disposal of medical waste. Diaz et al. [9] provided information on the quantities and properties of health-care wastes in various types of facilities located in developing countries, as well as in some industrialized countries. Shinee et al. [10] determined the current situation and characterization of health-care wastes generated in Ulaanbaatar, the capital of Mongolia. Lately, Mohamed et al. [11] analyzed HCWM practices including waste generation, segregation, storage, collection, transportation, treatment, and disposal in the Kingdom of Bahrain.

Although the abovementioned studies are useful in analyzing the current situation in developing countries, they fall short of addressing multiple subjective and conflicting criteria required to implement a comprehensive health-care waste management strategy. Alternatively, Brent et al. [12] have taken into consideration the multi-criteria nature of the HCWM problem. Brent et al. [12] integrated the analytic hierarchy process (AHP) with other systems approaches in order to establish primary HCWM systems that minimize infection risks in developing countries.

Evaluation of HCW treatment alternatives, which requires the consideration of multiple subjective and conflicting criteria with the involvement of a group of experts, is a highly important multi-criteria group decision making problem. In group decision making problems, consensus is an important indication of group agreement or reliability. Aggregation of expert opinions is critical for conducting an evaluation process effectively. The inherent imprecision and vagueness in criteria values concerning HCW treatment alternatives justify the use of fuzzy set theory.

This paper presents a hierarchical distance-based fuzzy multi-criteria group decision making framework for evaluating HCW treatment alternatives for Istanbul, one of the most crowded cities in Europe. As individuals intuitively attempt to be both as close as possible to the ideal and as distant as possible from the anti-ideal, the ideal and anti-ideal solutions are considered simultaneously in the proposed approach. The HCW treatment alternatives considered in this study include "incineration", "steam sterilization", "microwave", and "landfill". A hierarchy of evaluation criteria and their related sub-criteria is used to evaluate the HCW treatment alternatives.

The rest of the paper is organized as follows. The following section provides information on the major HCW treatment technologies. Section 3 presents the ordered weighted averaging (OWA) operator and the proposed framework for evaluation of the alternative waste management scenarios. The application of the proposed methodology to Istanbul's HCWM problem is presented in Section 4. Finally, conclusions are provided in Section 5.

2. Health-care waste treatment technologies

Waste generated from medical activities can result in negative impacts to public health and to the environment if inappropriate treatment and disposal are realized [13]. There exist four proven technologies for achieving significant pathogen destruction namely "incineration", "steam sterilization", "microwave", and "landfill".

Incineration, which is known as controlled-flame combustion to decline waste materials to noncombustible residue or ash and exhaust gases, is a remedial technology that destroys contaminants at high temperatures. Incineration is being used as the technology to dispose HCW generated by health-care institutions in Istanbul. It has been claimed as the most effective means for destroying infectious and toxic components, and for significantly reducing volume and weight [14]. However, the main inconvenience of medical waste incineration is the emission of

pollutants, some of them extremely toxic to the atmosphere.

Steam sterilization treatment combines moisture, heat and pressure to inactivate microorganisms. The factors that affect the efficacy of steam autoclave treatment of medical waste are those affecting the internal waste load temperature, steam penetration of the waste, and the duration of treatment [15].

Microwave disinfection is essentially a steam-based process, since disinfection occurs through the action of moist heat and stream generated by microwave energy. The factors which affect microwave treatment of medical waste include frequency and wavelength of the irradiation, duration of the exposure, destruction and moisture content of the waste material, process temperature, and mixing of the waste during treatment [15].

Sanitary landfilling is the preferred method of solid waste disposal in certain cases due to its low cost, minimal environmental impacts when designed and operated correctly, and effectiveness in controlling health risks. The primary objections to landfill disposal of hazardous HCW, especially untreated waste based on a perceived risk of release of pathogens to air and water [16].

3. Decision making framework

Selection of the best HCW treatment alternative for Istanbul involves the consideration of numerous performance attributes, yielding in general a multi-level hierarchical structure that enables to conduct a more effective analysis. Further, in general, crisp data are inadequate to model real-life situations. Since human judgments regarding preferences are often vague, it is difficult to estimate preference with an exact numerical value. A more realistic approach may be to use linguistic assessments rather than numerical values, that is, to suppose that the ratings and weights of the criteria in the problem are assessed by means of linguistic variables [17]. This paper proposes a robust fuzzy multi-criteria group decision making approach, which can address decision problems having multi-level hierarchical structure where a number of performance attributes are present.

In group decision making problems, aggregation of expert opinions is key to conducting the evaluation process appropriately. This paper employs ordered weighted averaging (OWA) operator, initially proposed by Yager [18], as the aggregation operator.

3.1. Aggregation operator

The OWA operator provides an aggregation which lies in between the "and" requiring all the criteria to be satisfied, and the "or" requiring at least one of the criteria to be satisfied. OWA operator differs from the classical weighted average in that coefficients are not associated directly with a particular attribute but rather to an ordered position. It encompasses several operators since it can implement different aggregation rules by changing the order weights.

Let $A = \{a_1, a_2, ..., a_n\}$ be a set of values to be aggregated, OWA operator F is defined as

$$F(a_1, a_2, ..., a_n) = \mathbf{w}\mathbf{b}^T = \sum_{j=1}^{n} w_j b_j, \qquad (1)$$

where $\mathbf{w} = \{w_1, w_2, ..., w_n\}$ is a weighting vector, such that $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$ and $\mathbf{b}$ is the associated ordered value vector, where $b_j \in \mathbf{b}$ is the j^{th} largest value in A.

A key characteristic of the OWA operator is the reordering of the arguments based upon their values, in particular an argument a_i is not associated with a specific weight w_i but rather a weight w_i is associated with a specific ordered position i of the arguments [19].

A crucial issue in employing the OWA operator for decision making is to determine its weights. In this study, the weights of the OWA

operator are computed using fuzzy linguistic quantifiers, which for a non-decreasing relative quantifier Q, are given by

$$w_i = Q(i/n) - Q((i-1)/n), \quad i = 1,...,n. \tag{2}$$

The non-decreasing relative quantifier, Q, is defined as [20]

$$Q(y) = \begin{cases} 0 & , y < a, \\ \dfrac{y-a}{b-a} & , a \le y \le b, \\ 1 & , y > b, \end{cases} \tag{3}$$

with $a, b, y \in [0,1]$, and $Q(y)$ indicating the degree to which the proportion y is compatible with the meaning of the quantifier it represents. Some non-decreasing relative quantifiers are identified by terms 'most', 'at least half', and 'as many as possible', with parameters (a,b) are $(0.3,0.8), (0,0.5)$, and $(0.5,1)$, respectively.

3.2. Hierarchical fuzzy multi-criteria group decision making approach

Fuzzy set theory, which was introduced by Zadeh [21] to deal with problems in which a source of vagueness is involved, has been utilized for incorporating imprecise data into the decision framework. A fuzzy set $\tilde{A}$ can be defined mathematically by a membership function $\mu_{\tilde{A}}(x)$, which assigns each element x in the universe of discourse X a real number in the interval [0,1].

A triangular fuzzy number $\tilde{A}$ can be defined by a triplet (a_1, a_2, a_3) with the membership function given as

$$\mu_{\tilde{A}}(x) = \begin{cases} \dfrac{x-a_1}{a_2-a_1}, & a_1 \le x \le a_2, \\ \dfrac{x-a_3}{a_2-a_3}, & a_2 \le x \le a_3, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Karsak [22] introduced a distance-based fuzzy multi-criteria decision making (MCDM) approach for technology selection that is based on the proximity to ideal solution concept and has the capability of incorporating both crisp and fuzzy data. The origins of Karsak's approach can be found in the multi-criteria decision aid named TOPSIS (technique for order preference by similarity to ideal solution) first developed by Hwang and Yoon [23].

This section presents the multi-expert version of the hierarchical distance-based fuzzy MCDM algorithm initially proposed by Karsak and Ahiska [24]. The proposed decision approach tackles the problem that an alternative with the shortest distance from the ideal may not have the farthest distance from the anti-ideal by taking into account the weighted distances from both the ideal and anti-ideal simultaneously. The stepwise representation of the proposed fuzzy MCDM algorithm is given below.

Step 1. Form a decision-makers' committee of z experts $(l = 1,2,...,z)$. Identify the alternatives, required selection criteria, and related sub-criteria in a hierarchical structure.

Step 2. Construct the decision matrices that denote the importance weight of criteria and related sub-criteria, and the fuzzy assessments corresponding to qualitative sub-criteria for each decision-maker.

Step 3. Compute the OWA weights for the decision-makers using Eqs. (2) and (3).

Step 4. Define the fuzzy value assigned to alternative i $(i = 1,2,...,m)$ with respect to sub-criterion k $(k = 1,2,...,p)$ of criterion j $(j = 1,2,...,n)$, importance weight of sub-criterion k of criterion j, and importance weight of criterion j for the l^{th} decision-maker as $\tilde{x}_{ijkl} = (x^1_{ijkl}, x^2_{ijkl}, x^3_{ijkl})$, $\tilde{w}_{jkl} = (w^1_{jkl}, w^2_{jkl}, w^3_{jkl})$ and $\tilde{w}_{jl} = (w^1_{jl}, w^2_{jl}, w^3_{jl})$, respectively. Calculate the aggregated fuzzy assessments of alternatives $(\tilde{x}_{ijk})$, the aggregated importance weights of sub-criteria $(\tilde{w}_{jk})$, and the

aggregated importance weights of criteria $(\tilde{w}_j)$, by utilizing Eq. (1). Consequently, the aggregated ratings of alternatives with respect to each sub-criterion can be calculated as $\tilde{x}_{ijk} = (x^1_{ijk}, x^2_{ijk}, x^3_{ijk})$, the aggregated importance weights of sub-criteria can be computed as $\tilde{w}_{jk} = (w^1_{jk}, w^2_{jk}, w^3_{jk})$, and the aggregated importance weights of criteria can be obtained as $\tilde{w}_j = (w^1_j, w^2_j, w^3_j)$.

Step 5. Normalize the aggregated decision matrix to obtain unit-free and comparable sub-criteria values. The normalized values for the data regarding benefit-related as well as cost-related sub-criteria are calculated via a linear scale transformation as

$$\tilde{r}_{ijk} = \begin{cases} \left(\dfrac{x^1_{ijk} - x^-_{jk}}{x^*_{jk} - x^-_{jk}}, \dfrac{x^2_{ijk} - x^-_{jk}}{x^*_{jk} - x^-_{jk}}, \dfrac{x^3_{ijk} - x^-_{jk}}{x^*_{jk} - x^-_{jk}} \right), & k \in CB_j \\ \left(\dfrac{x^*_{jk} - x^3_{ijk}}{x^*_{jk} - x^-_{jk}}, \dfrac{x^*_{jk} - x^2_{ijk}}{x^*_{jk} - x^-_{jk}}, \dfrac{x^*_{jk} - x^1_{ijk}}{x^*_{jk} - x^-_{jk}} \right), & k \in CC_j \end{cases} \quad (5)$$

where $\tilde{r}_{ijk}$ denotes the normalized value of $\tilde{x}_{ijk}$, m is the number of alternatives, n is the number of criteria, CB_j is the set of benefit-related sub-criteria of criterion j for which the greater the performance value the more its preference, CC_j is the set of cost-related sub-criteria of criterion j for which the greater the performance value the less its preference, $x^*_{jk} = \max_i x^3_{ijk}$ and $x^-_{jk} = \min_i x^1_{ijk}$.

Step 6. Aggregate the performance ratings of alternatives at the sub-criteria level to criteria level as follows:

$$\tilde{y}_{ij} = (y^1_{ij}, y^2_{ij}, y^3_{ij}) = \frac{\sum_{k=1}^{p} \tilde{w}_{jk} \otimes \tilde{r}_{ijk}}{\sum_{k=1}^{p} \tilde{w}_{jk}}, \forall i, j \quad (6)$$

where $\tilde{y}_{ij}$ represents the aggregate performance rating of alternative i with respect to criterion j and $\otimes$ is the fuzzy multiplication operator.

Step 7. Normalize the aggregate performance ratings at criteria level using a linear normalization procedure, which results in the best value to be equal to 1 and the worst one to be equal to 0, as follows:

$$\tilde{y}'_{ij} = \left(\frac{y^1_{ij} - y^-_j}{y^*_j - y^-_j}, \frac{y^2_{ij} - y^-_j}{y^*_j - y^-_j}, \frac{y^3_{ij} - y^-_j}{y^*_j - y^-_j} \right), \forall i, j \quad (7)$$

where $y^*_j = \max_i y^3_{ij}$, $y^-_j = \min_i y^1_{ij}$, and $\tilde{y}'_{ij}$ denotes the normalized aggregate performance rating of alternative i with respect to criterion j.

Step 8. Define the ideal solution $A^* = (r^*_1, r^*_2, \ldots, r^*_n)$ and the anti-ideal solution $A^- = (r^-_1, r^-_2, \ldots, r^-_n)$, where $r^*_j = (1,1,1)$ and $r^-_j = (0,0,0)$ for $j = 1, 2, \ldots, n$.

Step 9. Calculate the weighted distances from ideal solution and anti-ideal solution (D^*_i and D^-_i, respectively) for each alternative as follows:

$$D^*_i = \sum_{j=1}^{n} \frac{1}{2} \left\{ \max(w^1_j \left| y'^1_{ij} - 1 \right|, w^3_j \left| y'^3_{ij} - 1 \right|) + w^2_j \left| y'^2_{ij} - 1 \right| \right\}, \forall i \quad (8)$$

$$D^-_i = \sum_{j=1}^{n} \frac{1}{2} \left\{ \max(w^1_j \left| y'^1_{ij} - 0 \right|, w^3_j \left| y'^3_{ij} - 0 \right|) + w^2_j \left| y'^2_{ij} - 0 \right| \right\}, \forall i \quad (9)$$

Step 10. Compute the proximity of the alternatives to the ideal solution, P^*_i, by considering the distances from ideal and anti-ideal solutions as

$$P^*_i = D^-_i \Big/ \left(D^*_i + D^-_i \right), \forall i \quad (10)$$

Rank the alternatives according to P^*_i values in descending order, and identify the alternative with the highest P^*_i as the best alternative.

4. Application of the MCDM framework to health-care waste management in Istanbul

As a result of discussions with experts from Istanbul Metropolitan Municipality Environmental Protection and Waste Materials Valuation Industry and Trade Co. (ISTAC), capacity of alternative treatment technology is determined as 24 tons/day. We have defined four possible treatment technologies for the treatment of health-care wastes in Istanbul. Treatment systems for steam sterilization and microwaving are selected with pre-shredding component that exposes a greater surface area for treatment by utilizing a shredder that reduces the waste to a uniform and relatively small size matter. The considered alternatives are incineration (A_1), steam sterilization (A_2), microwave (A_3), and landfill (A_4).

Benefiting from the literature on the assessment of health-care treatment alternatives and discussions with the experts, economic criteria, environmental criteria, technical criteria, and social criteria, and their related sub-criteria are identified as the evaluation attributes in a hierarchical framework as depicted in Figure 1.

In this paper, the importance weights of various criteria and the ratings of qualitative criteria are represented using linguistic variables very low (VL), low (L), moderate (M), high (H) and very high (VH), which can be expressed by triangular fuzzy numbers as shown in Figure 2.

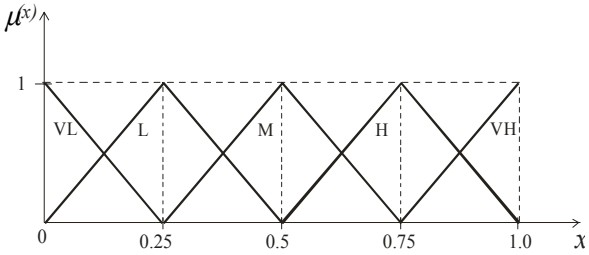


Figure 2. A linguistic term set where VL : (0, 0, 0.25), L : (0, 0.25, 0.5), M : (0.25, 0.5, 0.75), H : (0.5, 0.75, 1), VH : (0.75, 1, 1).

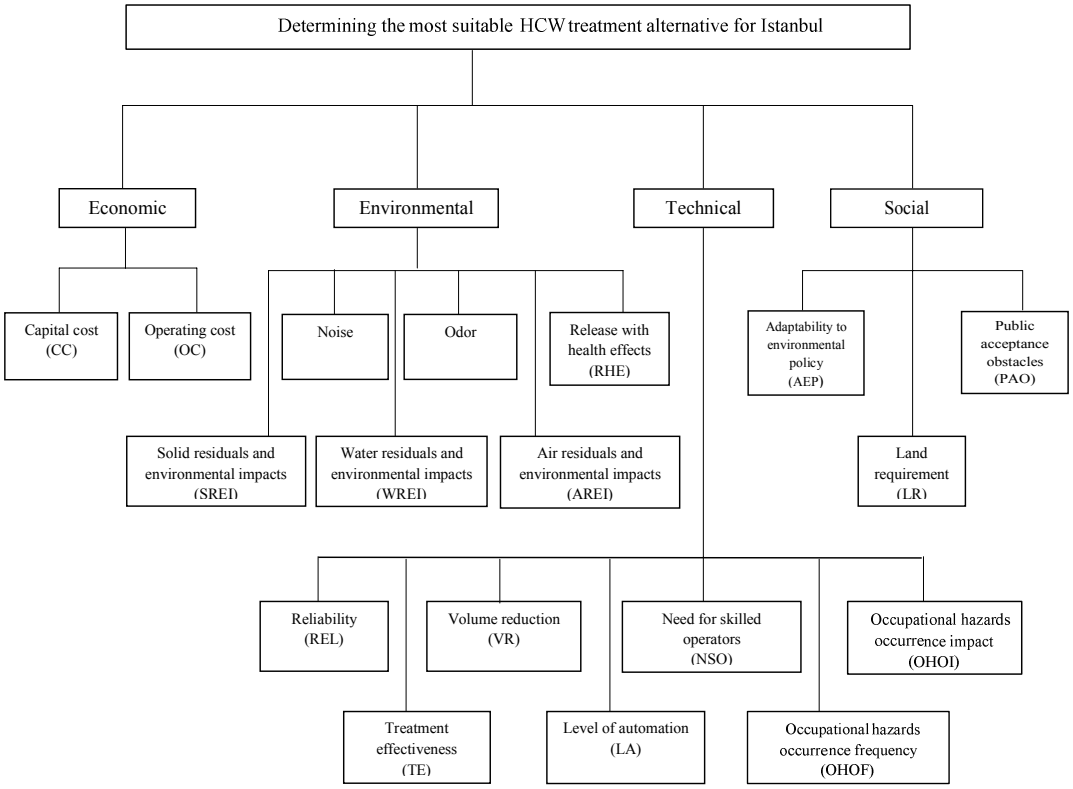


Figure 1. Hierarchical structure of the problem

The evaluation is conducted by a committee of five experts consisting of field experts from ISTAC, a university professor, and a technical advisor specialized in waste management. The computational procedure is summarized below.

Using the linguistic quantifier 'most' and Eqs. (2) and (3), the OWA weights for five decision-makers are computed as $\mathbf{w} = (0, 0.2, 0.4, 0.4, 0)$.

Then, the decision-makers' evaluations are aggregated using Eq. (1) to obtain the aggregated importance weights of criteria, the aggregated importance weights of sub-criteria, and the aggregated ratings of alternatives with respect to each sub-criterion. The results are presented in Tables 1 and 2, respectively.

In order to determine the HCW treatment alternatives' overall performance, the ideal and anti-ideal solutions are obtained as defined in step 8 of the hierarchical distance-based fuzzy MCDM algorithm.

Table 1. Aggregated importance weights of criteria and related sub-criteria

Criteria / Sub-Criteria	Importance weight
Economic	(0.500,0.750,1)
CC	(0.500,0.750,1)
OC	(0.750,1,1)
Environmental	(0.750,1,1)
SREI	(0.500,0.750,1)
WREI	(0.500,0.750,1)
AREI	(0.750,1,1)
Noise	(0,0.250,0.500)
Odor	(0.150,0.400,0.650)
RHE	(0.750,1,1)
Technical	(0.500,0.750,1)
REL	(0.500,0.750,1)
TE	(0.650,0.900,1)
VR	(0.150,0.400,0.650)
LA	(0.400,0.650,0.900)
NSO	(0.250,0.500,0.750)
OHOF	(0.500,0.750,1)
OHOI	(0.550,0.800,1)
Social	(0.300,0.550,0.800)
AEP	(0.650,0.900,1)
PAO	(0.400,0.650,0.900)
LR	(0.500,0.750,1)

Table 2. Aggregated ratings of alternatives with respect to sub-criteria

Sub-Criteria	A_1	A_2	A_3	A_4
CC	(0.750,1,1)	(0.250,0.500,0.750)	(0.250,0.500,0.750)	(0,0.250,0.500)
OC	(0.650,0.900,1)	(0.250,0.500,0.750)	(0.250,0.500,0.750)	(0,0.250,0.500)
SREI	(0.050,0.300,0.550)	(0.150,0.400,0.650)	(0.050,0.300,0.550)	(0.650,0.900,1)
WREI	(0.050,0.300,0.550)	(0,0.250,0.500)	(0,0.250,0.500)	(0.650,0.900,1)
AREI	(0.750,1,1)	(0,0.150,0.400)	(0,0.200,0.450)	(0.300,0.550,0.800)
Noise	(0.400,0.650,0.900)	(0.050,0.300,0.550)	(0.150,0.400,0.650)	(0.050,0.200,0.450)
Odor	(0.250,0.500,0.750)	(0.050,0.300,0.550)	(0.050,0.300,0.550)	(0.750,1,1)
RHE	(0.650,0.900,1)	(0.050,0.300,0.550)	(0,0.250,0.500)	(0.500,0.750,1)
REL	(0.650,0.900,1)	(0.400,0.650,0.900)	(0.250,0.500,0.750)	(0.300,0.550,0.800)
TE	(0.500,0.750,1)	(0.250,0.500,0.750)	(0.300,0.550,0.800)	(0,0.250,0.500)
VR	(0.450,0.700,0.900)	(0.050,0.300,0.550)	(0.050,0.300,0.550)	(0,0,0.250)
LA	(0.500,0.750,1)	(0.500,0.750,1)	(0.500,0.750,1)	(0,0.150,0.400)
NSO	(0.500,0.750,1)	(0.250,0.500,0.750)	(0.250,0.500,0.750)	(0,0.250,0.500)
OHOF	(0.150,0.400,0.650)	(0.150,0.400,0.650)	(0.050,0.300,0.550)	(0.200,0.450,0.700)
OHOI	(0.500,0.750,1)	(0.400,0.650,0.900)	(0.400,0.650,0.900)	(0,0.250,0.500)
AEP	(0.050,0.300,0.550)	(0.500,0.750,1)	(0.300,0.550,0.800)	(0,0.250,0.500)
PAO	(0.550,0.800,1)	(0,0.250,0.500)	(0.150,0.400,0.650)	(0.650,0.900,1)
LR	(0.400,0.650,0.900)	(0,0.250,0.500)	(0.050,0.300,0.550)	(0.750,1,1)

Subsequently, the weighted distances of each HCW treatment alternative from ideal and anti-ideal solutions, D_i^* and D_i^-, are calculated using Eqs. (8) and (9), respectively. At last, the proximity to the ideal solution for each HCW treatment is computed employing Eq. (10).

The results are illustrated in Table 3. Table 3 shows that "Steam sterilization", A_2, with the highest P_i^* value is the most preferred HCW treatment technology for Istanbul, and it is followed by "Microwave" (A_3). "Landfill" is positioned as the third while "Incineration" ranks as the last HCW disposal alternative mainly due to their unfavorable environmental and health impacts.

Table 3. Ranking of the HCW treatment alternatives

A_i	D_i^*	D_i^-	P_i^*	Ranking
A_1	2.076	1.684	0.448	4
A_2	1.268	2.673	0.678	1
A_3	1.309	2.600	0.665	2
A_4	1.969	1.834	0.482	3

5. Conclusions

Due to the recent upsurge in environmental problems caused by the HCW, determining an efficient HCWM system, which considers environmental, economic, social, and technical factors, is of outmost importance. Thus, especially in a large metropolis, choosing the adequate HCWM system appears as a multi-criteria decision making problem with a hierarchical nature. In this paper, a fuzzy multi-criteria group decision making framework is presented to rectify the problems encountered when using classical decision making methods in evaluating HCW treatment technologies. Besides having the capability of considering numerous attributes that are structured in a multi-level hierarchy, the proposed decision approach enables the decision-makers to use linguistic terms, and thus, reduces their cognitive burden in the evaluation process. This approach is apt to incorporate imprecise data represented as linguistic variables into the analysis. Weighted distances from both ideal and anti-ideal solutions are considered simultaneously in the presented approach.

According to the proposed fuzzy multi-criteria decision making approach, "Steam sterilization" is the best alternative treatment method for Istanbul since it minimizes the impact on the environment and demonstrates a commitment to public health. While "Landfill" is an economic alternative compared with other alternatives, it should only be used in a limited extent because of its several drawbacks for the environment and public health. "Incineration" ranks after non-incineration alternative technologies due to its high costs, and adverse environmental and health impacts.

For future research, extensions of the proposed methodology can be developed incorporating both subjective and objective weight assessments of the criteria and related sub-criteria. Further, a novice-friendly decision support system is to be developed for enabling the use of the proposed approach by users with limited skills in quantitative decision aids.

Acknowledgement

This research has been financially supported by Galatasaray University Research Fund.

6. References

[1] M. Kilic, “Optimization of the health-care waste handling and final disposal of the infectious wastes of the hospital-medical centers in the Anatolian Side of Istanbul”, M.S. Thesis, 2004, Bogazici University, Istanbul.

[2] B.K. Lee, M.J. Ellenbecker, and R. Moure-Ersaso, “Alternatives for treatment and disposal cost reduction of regulated medical wastes”, Waste Management, 2004, 24, pp. 143-151.

[3] B.A. Zeren, “Health-care waste management of the hospitals in the European Side of

Istanbul" M.S. Thesis, 2004, Bogazici University, Istanbul.
[4] L.F. Diaz, L.L. Eggerth, and G.M. Savage, "Alternatives for the treatment and disposal of healthcare wastes in developing countries", Waste Management, 2005, 25, pp. 626-637.
[5] A.Z. Alagoz, and G. Kocasoy, "Treatment and disposal alternatives for health-care waste in developing countries - a case study in Istanbul, Turkey", Waste Management & Research, 2007, 25, pp. 83-89.
[6] A.Z. Alagoz, and G. Kocasoy, "Determination of the best appropriate management methods for the health-care wastes in Istanbul", Waste Management, 2008, 28, pp.1227-1235.
[7] A.Z. Alagoz, and G. Kocasoy, "Improvement and modification of the routing system for the health-care waste collection and transportation in Istanbul", Waste Management, 2008, 28, pp. 1461-1471.
[8] M.E. Birpinar, M.S. Bilgili, and T. Erdogan, "Medical waste management in Turkey: A case study of Istanbul", Waste Management, 2009, 29, pp. 445-448.
[9] L.F. Diaz , L.L. Eggerth, Sh. Enkhtsetseg, and G.M. Savage, "Characteristics of healthcare wastes", Waste Management, 2008, 28, pp. 1219-1226.
[10] E. Shinee, E. Gombojav, A. Nishimura, N. Hamajima, and K. Ito, "Healthcare waste management in the capital city of Mongolia", Waste Management, 2008, 28, pp. 435-441.
[11] L.F. Mohamed, S.A. Ebrahim, and A.A. Al-Thukair, "Hazardous healthcare waste management in the Kingdom of Bahrain", Waste Management, 2009, 29, pp. 2404-2409.
[12] A.C. Brent, D.E.C. Rogers, T.S.M. Ramabitsa-Siimane, and M.B. Rohwer, "Application of the analytical hierarchy process to establish health care waste management systems that minimise infection risks in developing countries", European Journal of Operational Research, 2007, 181, pp. 403-424.
[13] W. Jangsawang, B. Fungtammasan, and S. Kerdsuwan, "Effects of operating parameters on the combustion of medical wastes in a controlled air incinerator", Energy Conversion and Management, 2005, 46, pp. 3137-3149.
[14] C.C. Lee, and G.L. Huffman, "Medical waste management / incineration", Journal of Hazardous Materials, 1996, 35, pp. 3495-3501.
[15] EPA, http://www.epa.gov/osw/nonhaz/industrial/medical/publications.htm., 2009.
[16] WHO, "Safe management of wastes from health-care activities", World Health Organization, Geneva, 1999.
[17] C.T. Chen, "Extension of the TOPSIS for group decision-making under fuzzy environment", Fuzzy Sets and Systems, 2000, 114, pp. 1-9.
[18] R.R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", IEEE Trans. Syst., Man, Cybern., 1988, 18 (1), pp. 183-190.
[19] D.P. Filev, and RR. Yager, "On the issue of obtaining OWA operator weights", Fuzzy Sets and Systems, 1998, 94, pp. 157–169.
[20] F. Herrera, E. Herrera-Viedma, and L. Martínez, "A fusion approach for managing multi-granularity linguistic term sets in decision making", Fuzzy Sets and Systems, 2000, 114(1), pp. 43-58.
[21] L.A. Zadeh, "Fuzzy sets", Information and Control, 1965, 8 (3), pp. 338-353.
[22] E.E. Karsak, "Distance-based fuzzy MCDM approach for evaluating flexible manufacturing system alternatives", International Journal of Production Research, 2002, 40(13), pp. 3167-3181.
[23] C.L. Hwang, and K. Yoon, "Multiple attribute decision making: methods and applications", Springer, Heidelberg, 1981.
[24] E.E. Karsak, and S.S. Ahiska, "Fuzzy multi-criteria decision making approach for transport projects evaluation in Istanbul", Lecture Notes in Computer Science, 2005, 3483, pp. 301-311.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Using Lego NXT Mindstorm Robots to Improve Learning in Computer Information Systems and Computer Science Courses at an Urban Community College

Michael Kaye and Yun Liu
Baltimore City Community College
mkaye@bccc.edu, yliu@bccc.edu

Abstract

An innovative approach to enhance the teaching and improve the learning in computer information systems and computer science courses has been implemented at Baltimore City Community College. Students in computer information systems and computer science courses are issued Lego NXT Mindstorm robots and learn to program them with RobotC, a programming language developed by the Carnegie Mellon University Robotics Academy. Using robotics applications to improve student learning in computer information systems and computer science courses at an urban community college is discussed in this paper.

1. Introduction

Intelligent professional service robotics is a growing world market. According to Junku Yuh, who leads the robotics program in the National Science Foundation's (NSF) Computer and Information Science and Engineering directorate, "Intelligent robots will be one of the engineering achievements of the 21st century" [1]. Further, the International Federation of Robotics predicts significant increases in robotics sales in the years 2011-2013 due to huge consumer markets opening up in China, India, Brazil, Russia, and the Middle East [2]. This robotics industry growth is now showing promise in areas ranging from the federal government to consumer markets.

In 2004, The Technology Collaborative conducted a study involving regional and national robotics companies, university researchers, and military and civilian contractors. The study identified robotics technician and technologist work-force needs that can be filled through an articulated robotics technology education program. Robotics technicians should be in high demand throughout the following industries: medical technology, military, homeland security, automotive, and aircraft. Knowledge that students gain from these industries is highly portable, qualifying graduates for openings in a variety of industries throughout the world [3].

To address the demand for robotics technicians, Baltimore City Community College (BCCC) applied for and was awarded an Advanced Technology Education grant from the NSF to develop a Robotics Technology Curriculum (RTC). The goals of the grant project are to: (a) educate underrepresented and qualified technicians for autonomous robotics and associated industries; (b) improve underrepresented students' attitudes about robotics technologies; and (c) produce underrepresented technician graduates with higher GPAs in current BCCC technician programs. The objectives of the grant project are to: (a) develop a Robotics Technician Curriculum at BCCC; (b) increase the success rate of the BCCC Electronics/Computer Information System/Computer Aided Drafting

Design technician programs through robotics instruction; (c) introduce robotics concepts to 11th and 12th graders in select high schools of the Baltimore City Public School System and improve their math problem solving skills through hands-on robotics exercises; (d) create a career pathway from the BCCC RTC to Morgan State University's School of Engineering for an engineering-related Bachelor's Degree; and (e) provide internship and job opportunities to BCCC RTC graduates. Efforts made towards achieving goal (c) and objective (b) are discussed in this paper.

2. BCCC technician programs

Currently, only 75% of BCCC students pass the core courses in the technicians' programs at BCCC. Two courses that serve as pre-requisites to the core courses in the technicians' programs at BCCC are CSC 108: Programming in C and CIS 118: Programming with Visual Basic. A pilot study was conducted with the CSC 108 and CIS 118 courses in which Lego NXT Mindstorm robots were used as an assistant or tool to teach students about programming principles. The purpose of the study was to determine if using robots as a virtual instructor enabled, mixed reality learning intervention increases the students' success rates in these courses. The use of robotics as an educational tool is growing in popularity and researchers believe this approach will revolutionize the way science, technology, engineering, and mathematics skills are taught [4, 5, 6]. The details and results of the pilot study are discussed in the following sections.

3. Pilot study

In the fall 2009 semester, two sections of CSC 108 and one section CIS 118 were offered at BCCC. One section of each course was chosen to participate in the pilot study while the second section of CIS 118 was used as a control group.

After students in the experimental group learned about the principles of for loops and functions in their respective classes, they were issued Lego NXT Mindstorm robots (see Figure 1), given a 60 minute introduction to the robot's hardware and RobotC, a C-based programming language developed by the Carnegie Mellon University Robotics Academy for Lego NXT Mindstorm robots, and issued the following challenge: program the robots to trace the perimeter of a two-foot square. The students were required to use one for loop and three functions in their program.

The desired solution to this assignment is to write a function that makes the robot drive forward, a function that makes the robot turn 90 degrees, and a function that makes the robot pause. These three functions can then be sequenced as follows in a for loop that iterates four times: drive forward, pause, turn 90 degrees, and pause. The RobotC code for this solution can be found in Section 6 of this paper.

When the students finished the assignment, they were given a 10-question quiz on the principles of for loops and functions. The quiz questions were conceptual in nature and multiple-choice in format. Each question was assigned a full point value of 10 points so the maximum number of points that could be earned on the quiz was 100. Students in the control group, who learned about the principles of for loops and functions in their classes, but were not given the hands-on robotics assignment, were also given this quiz. Sample questions from the quiz can be found in Section 7 of this paper.

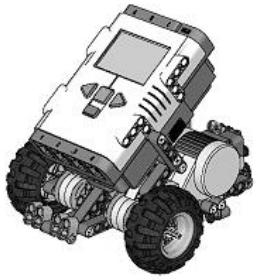


Figure 1. Lego NXT Mindstorm robot

4. Results

Seventeen students completed the robotics assignment and quiz. The mean score on the quiz for these students was 76.5 and the standard deviation was 16.6. Twelve students in the control group completed the quiz. The mean score on the quiz for the students in the control group was 57.9 and the standard deviation was 25.9.

A two sample t-test assuming unequal variances was done to test the null hypothesis which shows that there is no difference in the mean quiz scores for the two groups. The alternative hypothesis is that the mean score for the students who completed the robotics assignment is greater than the mean score for the control group. The one tail p-value for this test is 0.02.

It should be noted that an F-test was done to test for equal variances. The p-value of the F-test was 0.05, which is borderline significant. If a two sample t-test is done assuming equal variances, the one tail p-value is 0.01.

Regardless of whether the variances of the two samples are considered equal, the results of the pilot study are significant. The results suggest that using robots as a virtual instructor enabled, mixed reality learning intervention has increased the students' success rates in BCCC's CSC 108 and CIS 118 courses.

5. Next steps

A larger-scale study should be done to test the repeatability and scalability of the pilot study's results. Larger sample sizes should be used and robotics assignments should be given throughout the entire course, not just for two specific topics. Overall course grades could then be used to measure the effects of the robotics assignments. Other courses in the Electronics, Computer Aided Drafting Design, Computer Information Systems, Computer Science, and Computer Literacy areas should be included in the larger scale study.

6. Sample RobotC code

The following RobotC code could be used as a solution to the assignment that was given to the students in the pilot study.

```
void driveForward()
{
  // Drive forward at half power for
  // three seconds.
  motor[motorB] = 50;
  motor[motorC] = 50;
  wait1Msec(3000);
}

void turn90degrees()
{
  // Turn 90 degrees at half power.
  motor[motorB] = 50;
  motor[motorC] = 0;
  wait1Msec(250);
}

void pause()
{
  // Turn motors off for half of a second.
  motor[motorB] = 0;
  motor[motorC] = 0;
  wait1Msec(500);
}

task main()
{
  for(int i=0; i<4; i++)
  {
    driveForward();
    pause();
    turn90degrees();
    pause();
  }
}
```

7. Sample quiz questions

Students in the pilot study were given a 10 question quiz on the principles of for loops and functions. The quiz questions were conceptual in nature and multiple choice in format. A couple of questions are given below.

7.1 Sample question #1

A for loop would be most appropriate in which of the following situations?

(a) In a program that is used to make a machine hammer a nail 10 times,

(b) In a program that is used to make a machine drill a hole in a piece of wood until the machine operator hits the stop button

(c) In a program that is used to convert a Celsius temperature into a Fahrenheit temperature

7.2 Sample question #2

Which of the following is a reason why a programmer should consider using functions?

(a) The program becomes more complex when functions are used,

(b) They group sets of related statements together that may need to be executed several times

(c) They eliminate the need to use variables in the program

8. Acknowledgement

The authors wish to thank Professors Darius Kochesfahani and Warren Turner for their support of this project. This work was supported in part by an Advanced Technology Education grant from the National Science Foundation (DUE 0802454).

9. References

[1] A Special Report, Overview: The Quest of Intelligence, http://www.nsf.gov/news/special_reports/robotics/overview.jsp

[2] World Robotics, Executive Summary, 2010. http://www.worldrobotics.org/downloads/2010_Executive_Summary_rev.pdf.

[3] The Technology Collaborative: TTC Press Release: October 11, 2005, http://www.techcollaborative.org/default.aspx?id=press_Oct_11_2005.Date

[4] Miglino, Orazio, H. H. Lund, and M. Cardaci, Robotics as an Educational Tool, Journal of Interactive Learning Research, May 1998.

[5] Henrik Hautop Lund, and L. Pagliarini, Edutainment Robotics: Applying Modern AI Techniques, Proceedings of International Conference on Autonomous Minirobots for Research and Edutainment (AMIRE 2001).

[6] Lund, H.H., Modern Artificial Intelligence for Human-Robot Interaction, Proceedings of the IEEE Volume 92, Issue 11, Nov. 2004 Page(s): 1821 – 1838.

Quality Management of Road Safety Audits Process

Hesham S. Mahgoub
South Dakota State University
Hesham.mahgoub@sdstate.edu

ABSTRACT

Road Safety Audits have been used successfully worldwide for a number of years and are beginning to catch on in the United States. Road Safety Audit (RSA) is a proactive approach to improving highway safety. An RSA is an examination of an existing or proposed roadway by an independent and qualified team who prioritizes safety findings and reports on safety issues. The road safety audit process is best characterized as a proactive approach to road safety by addressing issues before accidents occur. The safety improvements resulting from RSAs can be achieved at a relatively low cost.
Clearly safety has to be a part of the routine maintenance activities. However, there is a need for an occasional assessment through the eyes of someone other than the usual maintenance and management staff, particularly at locations with significant crash history or the potential for crashes
This paper provides an overview of Quality Function Deployment (QFD) as a tool to determine priorities for enhancing Road Safety Audit (RSA) process to better satisfy client needs and requirements. Under QFD, RSA would be evaluated and characterised in regard to quality and operational efficiency.

INTRODUCTION

Road Safety Audits have been used successfully worldwide for a number of years and are beginning to catch on in the

United States. Road Safety Audit (RSA) is a proactive approach to improving highway safety. An RSA is an examination of an existing or proposed roadway by an independent and qualified team who prioritizes safety findings and reports on safety issues.

The RSA concepts were originally developed and introduced in the United Kingdom (UK) in 1989 and made mandatory by 1991. The benefits of such systematic checking were soon recognized around the world and many countries have since established their own similar systems. Through the 1990s, audits were introduced to other countries such as Australia, New Zealand and Canada. Audits have been conducted in the United States since the late 1990s. In the year 2000, Pennsylvania became the first state to formally adopt Road Safety Audits into its typical processes *(1)*.

RSA is a process concerned with the safety of all road users. It is a formal examination which applies safety principles from a multi-disciplinary perspective. The Federal Highway Administration defines RSA as a formal safety performance examination of an existing or future road or intersection by an independent audit team. It qualitatively estimates and reports on potential road safety issues and identifies opportunities for improvements in safety for all road users. *(2)*
A Road Safety Audit team must be an independent, qualified team which identifies and prioritizes safety findings and reports on safety improvement recommendations *(3)*. To achieve this independence, the freedom, ability, and comfort of team members to comment frankly on potentially controversial safety issues is crucial to the success of a Road Safety Audit. FHWA RSA guidance recommends that while a team member may be selected from within the local highway agency, this individual must be able to truly act independently *(4)*. To maintain this independence, the research approach was that the local road manager would not be a member of the team. In most of the case studies of any county road, a neighboring Highway Superintendent had been included in the RSA team.

In all cases, the teams included different types of expertise. Someone familiar with the road being reviewed (i.e. a school bus driver, a mail delivery person, a law enforcement officer, a road maintainer/blade operator, a truck driver, etc.), a multidisciplinary experience person who can bring synergy to the team effort, and at least one team member had professional experience in design, traffic operations, safety and was familiar with the design standards and the MUTCD *(5,6)*.

Quality Function Deployment (QFD) focuses on listening to the voice of the customer or client. It is a method that takes into account customer requirements at each stage of production or service delivery and process development by setting meaningful priorities and targeting activities relating to the service strategic goals.
Despite these diverse features of service in the public sector, QFD might still have a valuable role to play in future quality management initiatives due to its focus on integrating the voice of the customer with an understanding of the technical attributes of service delivery. This is achieved by translating attributes into a set of measurable parameters.

STUDY APPROACH

RSA projects were selected through the SDLTAP communications and commitments from South Dakota's local agencies *(7)*. The RSA team members were selected in each case study according to the nature and location of the project. All meetings and site visits for the RSAs in the case studies were conducted in one day.

The RSAs typically began with a start-up meeting attended by the Road Manager. Team members then reviewed the background information, crash data, and proposed plans furnished to them to gain insight into the road and identify any preliminary areas of safety concern. The team leader then described the RSA process. This included an overview of the RSA process with examples of safety issues that are typically encountered and mitigation measures to address them.

Following the start-up meeting, the RSA team conducted a field review. The purpose of the field review was to observe geometric and operating conditions for the roads. The RSA team observed road user characteristics (autos, trucks, pedestrians, agricultural vehicles, etc) and surrounding land uses. Figure 1 shows the study approach and Methodology.

THE USE OF QUALITY FUNCTION DEPLOYMENT (QFD) IN ROAD SAFETY AUDIT PROCESS

Step 1: identifications of customers' requirements

The first step of the QFD process is to identify the customer wants and needs to clearly identify what they expect from the RSA. Figure 2 describes the quality relates to customer expectations and not to the perception of those who provide the service. Of course, it is important to be clear about who the customers or clients are.

Step 2: obtaining customer importance ratings

In step 2, questionnaires and surveys are used to determine customer importance ratings. It is necessary to understand the relative importance that customers attribute to each requirement. The research team identified a list of safety issues to be considered when doing a road safety field review. It is not intended to be all inclusive, but used as the starting point. Figure 3 illustrates the main components of the review list. The review list asks a series of questions to stimulate thinking about possible safety issues. It is formatted as a checklist with space for notes to be taken during a review to identify specific safety issues for possible further consideration.

Step 3: identification of technology attributes

This step identifies the measurable and definable main features of the audit, including methods and processes necessary for its performing. Each of the main points of all findings were gathered and recorded. The team leader kept detailed notes of observations and a preliminary list of issues and proposed

recommendations for inclusion in the RSA report.

Step 4: construction of Safety Priority Evaluation Matrix

In this step, the team evaluated the risk and prioritized safety concerns and recommendations using historical crash records and the concept of risk. A RSA Safety Priority Evaluation Matrix form had been used to assess risk based on the likelihood of an event and its possible consequences. The purpose of this form is for the RSA team members to have discussion on the reasons they have for identifying a safety issue as a risk and lead to a team consensus on the highest priority recommendations. Figure 4 shows the priority evaluation matrix.

Step 5: feasibility study

This step is used to identify the safety features that can be realistically improved, given a set of constraints. Customer importance ratings represent the areas of greatest interest and highest expectations as expressed by the customer. The opportunities for improvement identified in this way can then be linked to both the RSA and, more broadly, to the agency's strategic vision.

Following the on-site portion of the RSA, the team wrote and issued the RSA report. The report included the summary of the field review, identified and prioritized safety issues, risks, and recommendations. The team leader drafted the report and provided an opportunity for each team member to review and comment. Every effort was made to complete the formal report within a relatively short time frame (two weeks).

The Road Managers were encouraged to write a brief response letter after reviewing the final report and recommendations. The research team encouraged them to outline what actions will be taken to each safety concern listed in the RSA report. The Road Manager had the opportunity to agree or disagree with the recommendations. If there was disagreement the response "no action will be taken" was documented.

Step 6: select quality elements requirements to be deployed in the reminder of the process

From the analysis of the 13 RSA visits, it can be seen that the ratings for the importance of quality elements include agency support and willingness to implement findings. The use of small, independent, and knowledgeable teams that consists of 3-5 members with expertise in design, safety, signing, law enforcement, construction, and maintenance has been proven successful.

The team looks into a variety of roadway features that are carefully inspected. At the conclusion of the visit, the team comes up with recommendations for improvements to roadway width, surface condition, pavement/shoulder drop offs, short pipe/box culverts, signing and delineation deficiencies, unprotected obstacles within the clear zone, obstructions in the ROW, and sight distance issues, to name a few.

Step 7: quality planning and implementation

This step is used to support the RSA process improvements, testing and evaluation and to translate outcomes into measurable terms. This step addresses the need to go beyond the identification of safety countermeasures priorities to establish realistic performance targets, and identify how the RSA can best achieve the highlighted goals. A framework for the implementation of QFD approach within a representative highway agency is shown in Figure 5.

CONCLUSION

QFD is an innovative, systematic methodology and a valuable tool for managing the processing of road safety audit. Through QFD's focus on listening to the voice of the customer; quality can be built into the audit process at all stages. Even with the budget restraints faced by most local governments, there are many low cost safety countermeasures available to address the most common and predictable crash types: run off the road rollovers and striking fixed objects. To make improvements, it is often useful to

think about safety from a new perspective or point of view. Recognizing that most local road managers are faced with the ever increasing challenges of keeping up with maintenance needs while being faced with difficult resource constraints, it can be productive to step back and think about safety of the roads from the perspective of those who use our roads. A fresh look can identify safety improvements that may be accomplished at low cost.

By applying QFD approach to the RSA, an agency can improve safety and demonstrate how it is taking action to reduce crashes. A number of benefits of Road Safety Audits have been identified, including:

- Demonstrate a proactive approach to safety (i.e., the agency does not have to wait until accident history identifies a problem)
- Provide an independent, unbiased perspective of safety issues and opportunities for improvement by involving outside expertise in the evaluation process.
- Identify low cost safety/high value improvement opportunities
- Promote awareness of safe design and maintenance practices
- Potentially reduce costs by identifying safety issues and correcting them before projects are built
- The RSA can be used as the basis for making a request for special safety funding

REFERENCES

1. Wilson E. M., *NCHRP Synthesis of Highway Practice 321: Roadway Safety Tools for Local Agencies*, Transportation Research Board, National Research Council, Washington, D.C., 2003.
2. FHWA, "Road Safety Audit Guidelines", U.S. Department of Transportation, Federal Highway Administration, Office of Safety, Publication No. FHWA SA-06-06, 2006
3. Austroads. *Road safety audit*. Austroads National Office, Austrália, 1994.
4. Wilson, F.R. and Hildebrand, E.D., *Road Safety Audits for New Brunswick*, Presentation to New Brunswick Department of Transportation, Fredericton, New Brunswick, April 9, 1999
5. *Manual on Uniform Traffic Control Devices*, Millennium Edition, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2000.
6. FHWA, "ROAD SAFETY FUNDAMENTALS", U.S. Department of Transportation, Federal Highway Administration, Office of Safety, Publication No. FHWA SA-05-011, September 2005
7. .Mahgoub, H, Marshall, R. Selim, A., and Skorseth, K. "Local Rural Road Safety Audit Guidelines and Case Studies" in Transportation Research Record, Washington, D.C. 2010.

Figure 1: QFD Approach to the Road Safety Audit Process

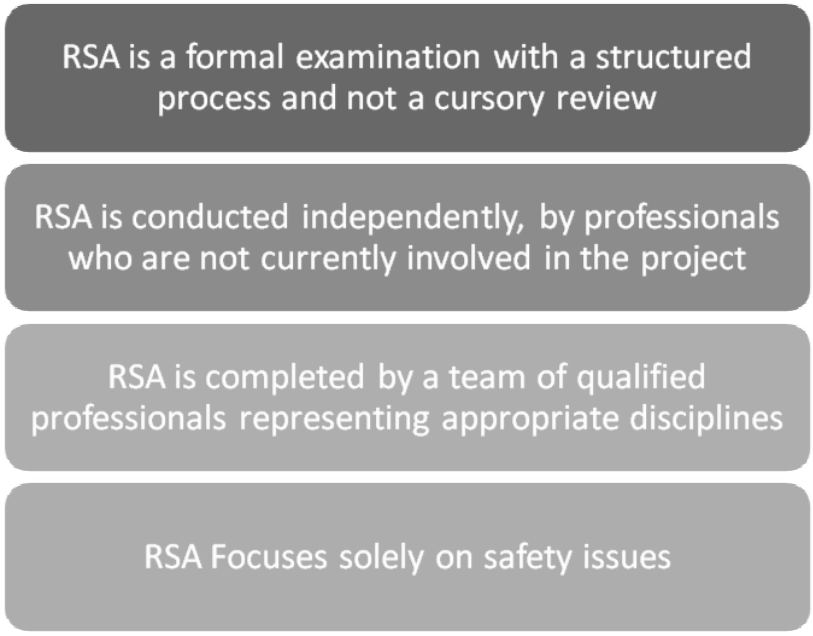


Figure 2: Quality Relates to Customer Expectations to the Road Safety Audit Process

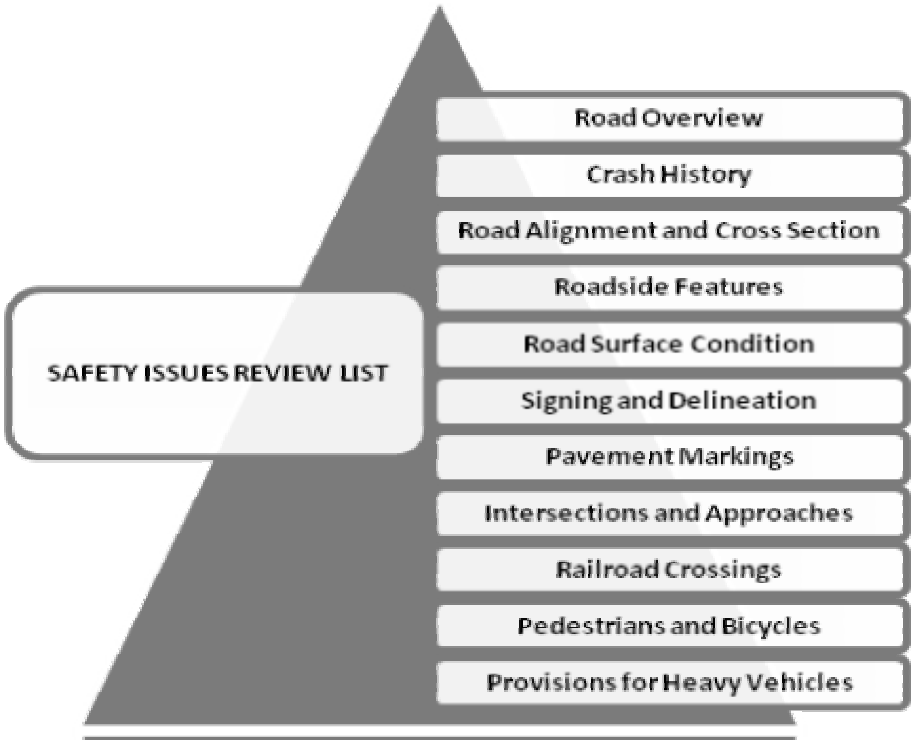


Figure 3 Safety Issues Review List

Safety Priority Evaluation Matrix

Description of issue/hazard:
Location:

	Very Low	Low	Medium	High	Very High
Exposure					
Probability					
Consequence					

Comment on Safety Risk:

Recommendation

Figure 4 RSA Safety Priority Evaluation Matrix

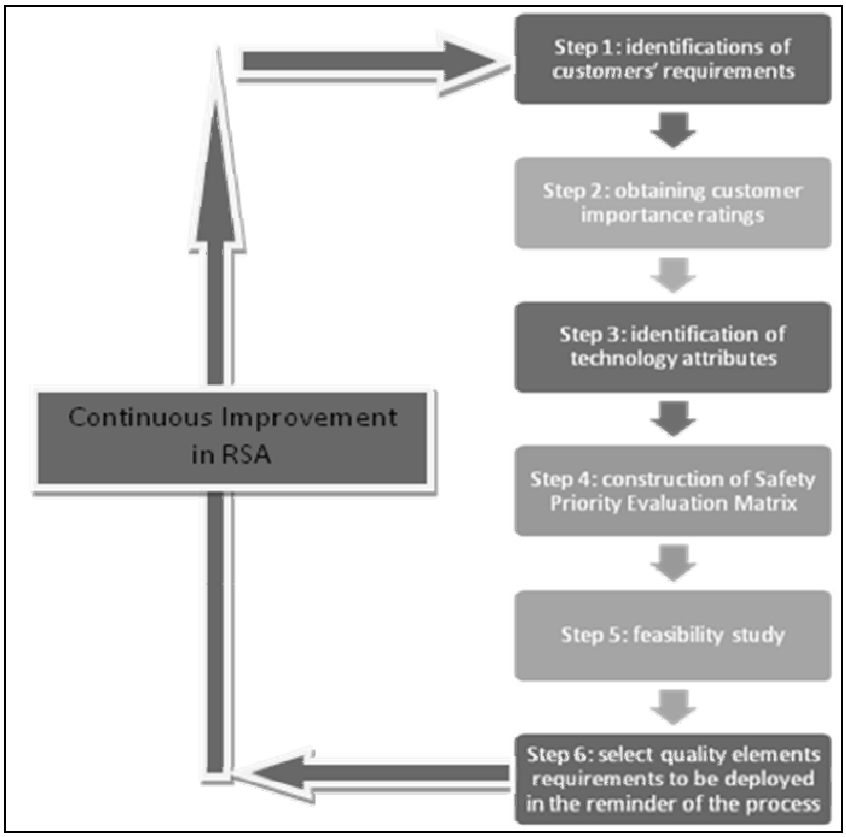


Figure 5 QFD Approach within the Highway Agency

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Reliability Analysis on a Public Transportation Bus Motor Failures

R. Radharamanan
Mercer University
radharaman_r@mercer.edu

Angela P. Ansuj
Federal University of Santa Maria, Brazil
angelaansuj@yahoo.com

Abstract

This paper analyzes the failures of bus motors for a large public transportation company. Both parametric and nonparametric techniques were used to determine the appropriate distribution governing these failures. The first step in the analysis was to plot the histogram, reliability function, and hazard function. After the distribution was visualized, six different parametric distributions were plotted and the coefficients of determination were calculated. Using these values, the best fit distribution was selected. The parameters such as failure rate (λ), mean time to failure (MTTF), and the confidence intervals for the MTTF were also found for each of the failures. The results were then compared to published data available in the literature. It was concluded that the Normal distribution can be used as an appropriate estimate for the first motor failures; the second motor failure can be viewed as a combination of both the Normal and Exponential distributions; the Exponential distribution can be used for the subsequent failures.

1. Introduction

The expense of engine spare parts and maintenance is steadily increasing in today's competitive market and public transportation spare parts suppliers must find effective ways to reduce cost. It is critical for them to understand how an engine will age due to its usage (number of miles performed). There are typically two ways to model the reliability of complex machinery. The first approach is to view it from a micro level, taking the reliability of each part into account. Since an engine has over 300 parts, this becomes very difficult and labor intensive. In this case, it is so time consuming that results may not be available when needed and may not be cost effective. The second approach is to view it from a macro level considering the failure of the whole engine [1-3].

Theoretical models on reliability, accelerated life testing, risk analysis, survival analysis, and safety assessment are presented and discussed by several authors [4-9]. Ahmadi et al [10] described methodology for the identification of different operational consequences and associated costs, cause by known or suspected technical difficulties or malfunctions caused by aircraft system failure, in order to facilitate and enhance the capability of taking correct and efficient decisions for the design of required maintenance tasks, effective cost analysis, and possible design modifications.

Failure analysis on vehicle engine crankshaft was carried out by several authors [11-13]. Xu et al [12] studied the causes of failure of a truck diesel engine crankshaft and reported that fatigue is the dominant failure mechanism of the crankshaft. Fracture failure analysis of ductile cast iron crankshaft in a vehicle engine indicated that the failure mechanism of the crankshaft was

fatigue fracture resulting from co-effect of bending and twisting [13]. Forecasting of engine removal was analyzed using neural network and hazard function models [14].

Davis [1], in his report on "An Analysis of Some Failure Data", determined the distributions governing bus motor failures for a large city bus company. Using the chi-squared method, he found that the majority of the failures fit an Exponential distribution. Parametric multiplicative intensities models were fitted to bus motor failure data [2]. Reliability modeling and aging characteristics of bus motor were studied using four different competing risk models each with three parameters [3]. The objective of this paper is to use the coefficient of determination to find the best distribution for the bus motor failure data and to compare the results with the published results in [1].

2. Methodology

The engine failure data for 191 buses were obtained from a large city bus company. Failures occurred when the motor was unable to run due to a broken part or when the maximum power produced could not exceed a specified percentage of the normal value. Failures of small parts that could easily be replaced were not included in the data collection and analysis. The data collected for the number of miles to the motors' first, second, third, fourth, and fifth failures were grouped in 20,000 mile intervals. Table 1 shows the complete set of the grouped data.

Initially, nonparametric methods were used in order to visualize the behavior of the data. These methods also help in choosing the type of distribution that would be appropriate for the given data. These methods include: plotting a histogram, reliability function, and the hazard function [4-7]. For grouped data, the reliability is calculated using:

$$\hat{R}(t_i) = \frac{n_i}{N}, \quad i = 1,2,.....,M \tag{1}$$

where: n_i is the number of motors that did not fail, N is the total number of motors, and M is the time intervals in which the data are grouped. The hazard function [4-5] is plotted using,

$$\hat{H}(t_i) = \ln N - \ln n_i \tag{2}$$

Table 1. Bus motor failures [1]

Initial Failure	
Distance Interval, Thousands of Miles	Observed # of Failures
0-20	6
20-40	11
40-60	16
60-80	25
80-100	34
100-120	46
120-140	33
140-160	16
160-180	2
180-up	2
Total	191

Second Failure	
Distance Interval, Thousands of Miles	Observed # of Failures
0-20	19
20-40	13
40-60	13
60-80	15
80-100	15
100-120	18
120-160	7
160-up	4
Total	104

Third Failure	
Distance Interval, Thousands of Miles	Observed # of Failures
0-20	27
20-40	16
40-60	18
60-80	13
80-100	11
100-up	16
Total	101

Fourth Failure	
Distance Interval, Thousands of Miles	Observed # of Failures
0-20	34
20-40	20
40-60	15
60-80	15
80-up	12
Total	96

Fifth Failure	
Distance Interval, Thousands of Miles	Observed # of Failures
0-20	29
20-40	27
40-60	14
60-80	8
80-up	7
Total	85

Six different distributions were used to determine the behavior of the bus motor failure and to fit the best approximated curve for the data collected. These were the Exponential, Normal, Lognormal, Weibull, Maximum Extreme Value, and Minimum Extreme Value distribution curves. To determine how well they fit the data, their corresponding probability curves were plotted. Using the least squares fit, the best fit line was plotted and the corresponding coefficient of determination, R^2 value was determined for each of the distributions. For the best fit distribution, its R^2 value should be closest to one. Table 2 shows the probability graphing information for different distributions.

Table 2. Probability graphing information [2]

Distribution	F(t)
Exponential	$1-e^{-\frac{t}{\theta}}$
Normal	$\Phi\left(\frac{t-\mu}{\sigma}\right)$
Lognormal	$\Phi\left(\frac{1}{\omega}\ln\frac{t}{t_0}\right)$
Weibull	$1-e^{-\left(\frac{t}{\Theta}\right)^m}$
Max. Extreme Value	$\exp\left(-e^{-\frac{t-\mu}{\Theta}}\right)$
Min. Extreme Value	$1-\exp\left(-e^{-\frac{t-\mu}{\Theta}}\right)$

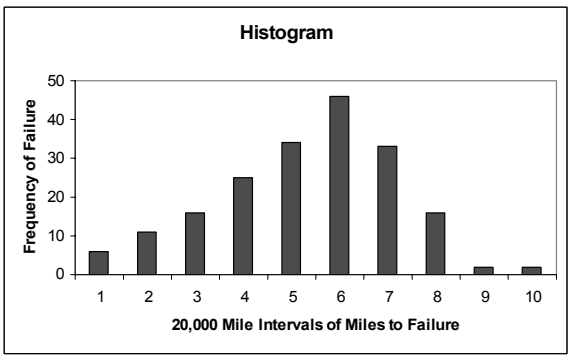
y(F)	x(t)	Parameters	
$\ln\left(\frac{1}{1-F}\right)$	t	0	θ
$\Phi^{-1}(F)$	t	μ	σ
$\Phi^{-1}(F)$	$\ln(t)$	t_0	ω
$\ln\left(\ln\left(\frac{1}{1-F}\right)\right)$	$\ln(t)$	θ	$\frac{1}{m}$
$-\ln\left(\ln\left(\frac{1}{F}\right)\right)$	t	Θ	μ
$\ln\left(\ln\left(\frac{1}{1-F}\right)\right)$	t	Θ	μ

The Exponential distribution is generally used for constant failure rate models. The Normal distribution, commonly referred to as Gaussian distribution is the most used distribution in statistics. In reliability, they are used for time-dependent failure rate models.

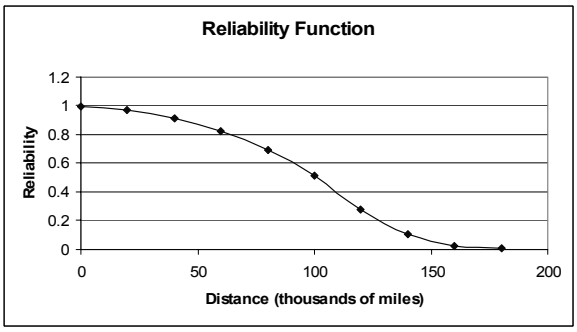
Although the Normal distribution fits a host of phenomenon, it is generally not a reasonable representation for experimental data. Lognormal distributions are usually used to describe failure caused by fatigue, uncertainties in failure rates, and other phenomenon. The Weibull distribution is widely used for brittle materials in modeling their distribution of times to failure and of their strengths. It works well for situations where the largest of many flaws is responsible for failure. Extreme value distributions are good in situations where the number of variables or flaws is very large. The Minimum Extreme Value distribution is used as an alternative to the Weibull distribution. When several loads are applied to the system, the Maximum Extreme Value distribution works well when the system's failure depends on the maximum load.

3. Results and discussions

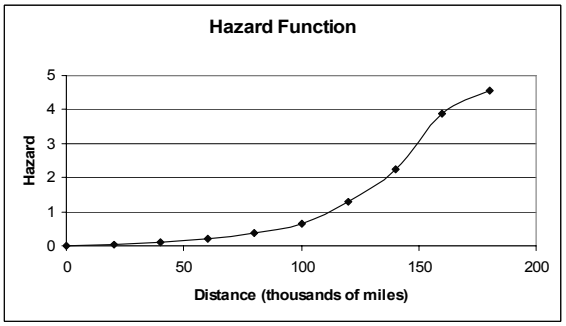
The nonparametric plots (histogram, reliability function, and hazard function) for the initial bus motor failures are shown in Figure 1. The histogram in Figure 1(a) suggests that the data could possibly follow a Normal distribution. The reliability is defined as the probability that the system will function properly for a specified amount of time. Figure 1(b) shows that the reliability of the bus motors decreases with increasing distance travelled. The hazard function plot in Figure 1(c) is concave upward. This means that the failure rate increases with time or, in this case, distance.



(a)



(b)

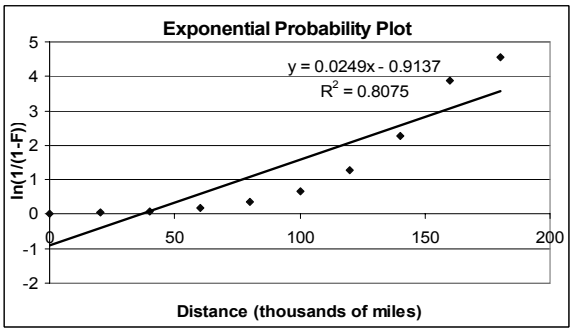


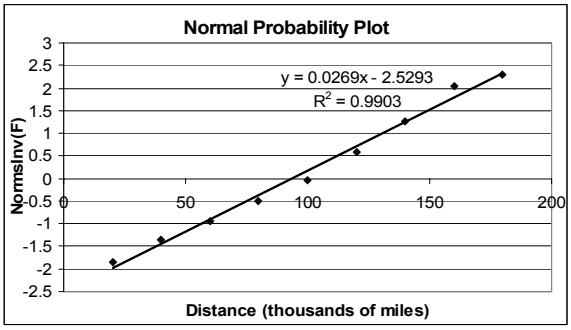
(c)

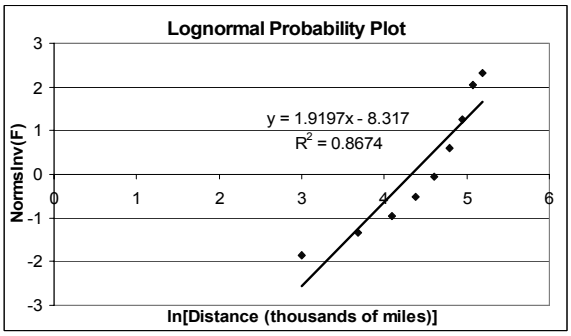
Figure 1. Nonparametric plots for the initial bus motor failures

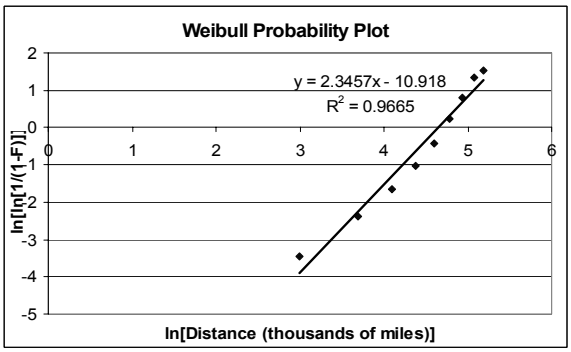
These plots for the second, third, fourth, and fifth motor failures were plotted in the same manner. The hazard function plot for the second motor failures was almost linear having a slight curvature that was concave upward. In this case, the failure rate increases slightly with distance travelled. For the remaining motor failures, the hazard functions were linear suggesting that the failure rates remain constant. The six probability plots for the initial bus motor failures are shown in Figure 2.

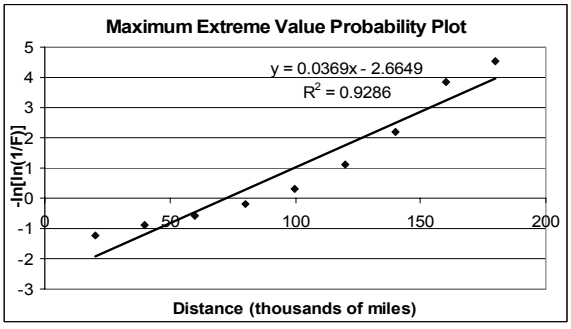
Based on its R^2 value of 0.9903, the Normal distribution was found to fit the data the best. Figure 3 shows the cumulative distribution function plots for the six distributions. For the distribution to be a close approximation to the data, its plotted line should follow the cumulative distribution function of the initial motor failure data indicated by the X's.











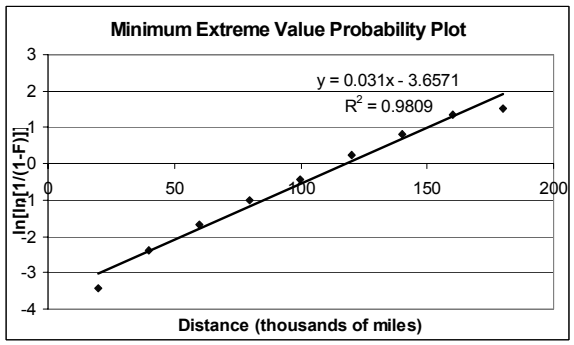


Figure 2. Probability plots for the initial bus motor failures

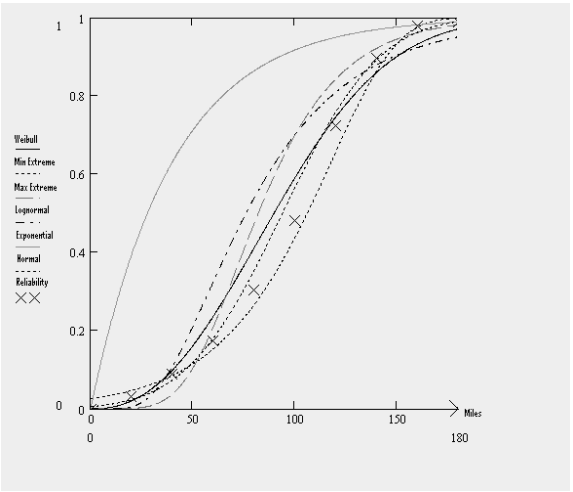


Figure 3. Distributions plotted with cumulative distribution function

The plots show that all of the distributions except the Exponential are very similar in the interval from 0 to 50,000 miles and from 150,000 miles and above. This is because these distributions follow a similar form. The Normal and Lognormal distributions are related and in the same way the Weibull and Extreme Value distributions are related. Each one of them is variations of the bell curve. These plot show that the cumulative distribution function of the first motor failure data follows the cumulative Normal distribution curve the best. This confirms that the Normal distribution could be used to estimate the data. This distribution is reasonable since each motor starts in relatively the same condition. The bulk of the motors will fail around the mean value with a few failing at various intervals from the mean.

The linear regression of the Normal probability plot indicates that the mean was 94,026 miles and the standard deviation was 37,175 miles. The mean time to failure [4-5] was approximated to be 96,702 miles using the expression:

$$\hat{\mu} = \sum_{i=1}^{M} \bar{t}_i f_i \Delta_i \qquad (3)$$

where: $\bar{t}_i = \frac{1}{2}(t_{i-1} + t_i)$; $f_i = \frac{n_{i-1} - n_i}{N\Delta_i}$; and $\Delta_i = t_i - t_{i-1}$. The failure rate for this data increases with time and is expressed as:

$$\lambda(t) = 4\times10^{-6} \exp\left[-\frac{(t-96702)^2}{2.76\times10^9}\right]\left[1-\Phi\left(\frac{t-96702}{37175}\right)\right]^{-1} \qquad (4)$$

For a 90% confidence interval μ ranges from 84,614 to 108,789 miles.

For the second, third, and fourth motor failures, the Normal distribution was found to be the best. The Normal probability plot for the second bus motor failures with R^2 value of 0.9897 is shown in Figure 4. The values of mean, standard deviation, mean time to failure, and the confidence intervals for the second, third, and the fourth motor failures are given in Table 3. The mean values are found to be decreasing because they will be more likely to fail earlier after the repair due to the increased mileages.

This seems out of the ordinary since, at this point, the data should start to follow an Exponential distribution because the motors have combinations of new and old parts. Also, according to the hazard function plots, the failure rates for these motor failures were found to be constant. In these cases, Exponential distributions are generally used. The data were also censored at the last interval which could have skewed the plot.

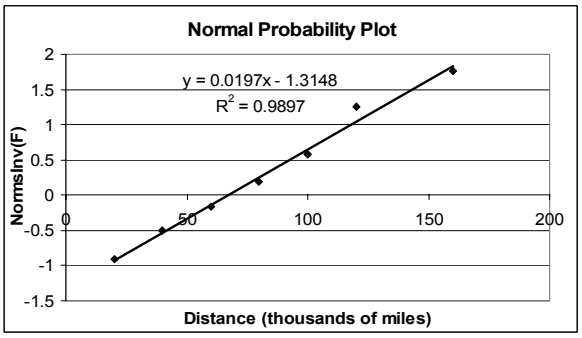


Figure 4. Normal probability plot of second bus motor failures

Table 3. Parameters for second, third and fourth bus motor failures (Normal distribution)

Failure	Mean from Excel	Standard Deviation	MTTF	Confidence Interval
Second	66741	50761	69904	58999<μ<83011
Third	49266	49261	52574	45740<μ<63089
Fourth	34824	40000	39792	33823<μ<48944

Even though the R^2 values indicate a Normal distribution for the third and fourth motor failures, the histograms suggest that the data more closely follow an Exponential distribution. For this reason, the Exponential distribution was chosen for these failures. The Exponential distribution parameters for these failures are listed in Table 4.

Table 4. Parameters for third and fourth bus motor failures (Exponential distribution)

Failure	θ	Failure Rate	MTTF	Confidence Interval
Third	55556	0.019	52574	45740<μ<63089
Fourth	-40000	0.025	39792	33823<μ<48944

The distribution for the second motor failures was the most difficult to determine. Figure 5 shows the histogram of these failures.

As can be seen from the histogram, the data did not follow one trend, but a combination of trends. The failures start off high and then level out before they increase and decrease a second time. Because of this behavior, it seemed to follow a combination of both the Normal and

Exponential distributions. The hazard function plot indicated that the failure rate is constant. This rate was calculated and found to be 0.0143.

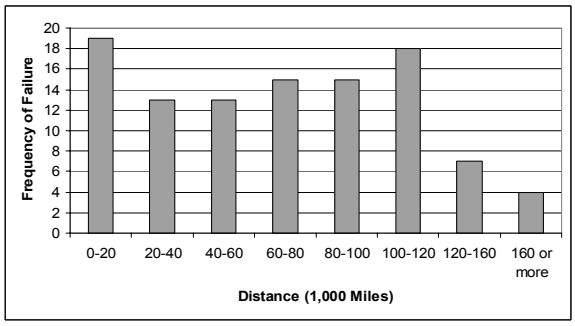


Figure 5. Histogram of second bus motor failures

The Weibull distribution fits the fifth bus motor failures the best. The shape and scale parameter were found to be 1,288 and 38,914 respectively. This distribution fits well since by the time the motor fails for the fifth time the failure will be due to major component wear. The failure rate was 0.028 with a MTTF of 35,177 miles and a confidence interval ranging from 30,252 to 43,619 miles. However, like the third and fourth motor failures, the Exponential distribution better fits the histogram.

These results differ from the study done by Davis [1]. While he did find that the first motor failures followed a Normal distribution, he found that the third, fourth, and fifth motor failures followed an Exponential distribution. In his analysis, he used a chi-squared index to determine whether the motor failures fit the Normal distribution or Exponential distribution [2-5]. The chi-squared index was calculated using the expression:

$$\chi^2 = \frac{(F_o - F_e)^2}{F_e} \tag{5}$$

where F_o is the observed frequency and F_e is expressed as:

$$F_e = \left(\frac{N}{\overline{T}}\right) e^{-\frac{t}{\overline{T}}} \tag{6}$$

where: $\overline{T}$ is the sample estimate of mean distance to failure. To determine whether the distribution fits the data set, the calculated chi-squared index is compared to the critical chi-squared value found in the table by using the degrees of freedom and the probability error threshold for the data set. If the calculated value is larger than the critical value, the data relationship is statistically significant. This difference in the analysis was probably the cause for the discrepancies in the distribution choice. It is also important to note that, although the analysis carried out in this paper showed the Normal distribution to be the best fit for the second, third, and fourth motor failures, the Exponential distribution better represents the histograms.

Davis' analysis was also performed on more detailed data sets [1]. His histograms show that the data he used were grouped in 10,000 mile intervals whereas the data he supplied in the Tables (Table 1) and that were used in this paper were grouped in 20,000 mile intervals. Furthermore, some crucial information is missing in the data set. For instance, all 191 buses had an initial bus motor failure; however, some of these buses did not have a second, third, fourth, or fifth failure. It is unclear why these buses were censored out of the data. It is possible that they either never had a second failure or repairs were not possible after the first failure. Furthermore, the sequences in which the failures occurred were not properly recorded. This limits the depth of analysis that can be performed on the data sets.

4. Conclusions

Using both parametric and nonparametric distributions, it was found that the first through fourth motor failures could be modeled using the

Normal distribution. The fifth motor failures were found to follow the Weibull distribution. However, it was found that the third through fifth failures can also be modeled with the Exponential distribution. Table 5 shows the characteristics for each of these failures.

Table 5. Characteristics of each bus motor failure

Failure	Distribution	Failure Rate	MTTF	Confidence Interval
First	Normal	Equation 4	96702	$84614<\mu<108789$
Second	Exponential/ Normal	0.0143	69904	$58999<\mu<83011$
Third	Exponential	0.019	52574	$45740<\mu<63089$
Fourth	Exponential	0.025	39792	$33823<\mu<48944$
Fifth	Exponential	0.028	35177	$30252<\mu<43619$

When compared to the published results from the literature [1], the second through fifth motor failures differed in that the author found that the Exponential distribution modeled these failures more accurately. The differences in the analyses were more likely caused by the different methods of goodness of fit. While in this paper, the coefficient of determination was used to determine the distribution that best fit the data analyzed, the author [1] used the chi-squared test. The author also did not supply the full set of data used in his analysis or the details of the dependencies of the failures to one another. From the analysis, it can be suggested that the Normal distribution can be used to determine the reliability of the first motor failures while the Exponential distribution can be used for each subsequent failures. The second motor failures can be viewed as a combination of both the Normal and Exponential distributions.

5. References

[1] D. J. Davis, "An Analysis of Some Failure Data", J. Am. Statist. Ass. V. 47, 1952, pp. 113-150.

[2] J. K. Lindsey, "Parametric Multiplicative Intensities Models Fitted to Bus Motor Failure Data", Appl. Statist., V. 46, No. 2, 1997, pp.245-252.

[3] Jiang, R., Reliability Modeling and Aging Characteristics of Bus-Motor, International Journal of Performability Engineering, Volume 6, Number 3, pp. 279-287, May 2010.

[4] Lewis, E.E., Introduction to Reliability Engineering, John Wiley & Sons, Inc. Canada, 1994.

[5] Zacks, S., Introduction to Reliability Analysis, Springer-Verlag, Inc., New York, NY p.135, 1992.

[6] Hoyland, A., and M. Rausand, System Reliability Theory: Models and Statistical Methods, John Wiley & Sons, Inc. New York, NY, p. 275 – 294, 2009.

[7] Modarres, M., M. Kaminskiy and V. Krivtsov, Reliability Engineering and Risk Analysis: A Practical Guide, 2nd Edition, Taylor & Francis, Inc., 2009.

[8] Huber, C., N. Limnios, M. Mesbath and M. Nikulin, Mathematical Methods in Survial Analysis, Reliability and Quality of Life, (E-Book), John Wiley and Sons, Inc., 2010.

[9] Thomson, J. R., Engineering Safety Assessment, An Introduction, John Wiley and Sons, Inc., New York, NY, p. 5, 1987.

[10] Ahmadi, A., U. Kumar and P, Soderholm, Operational Risk of Aircraft System Failure, International Journal of Performability Engineering, Volume 6, Number 2, pp. 149-158, March 2010.

[11] Alfares, M. A., A. H. Falah and A. H. Elkholy, Failure Analysis of a Vehicle Engine Crankshaft, J. Fail. Anal. and Prven., 7, pp. 12-17, 2007.

[12] Xu, X., X. Yu and Z. Yang, Truck Diesel Engine Crankshaft Failure Analysis, J. Fail. Anal. and Preven., Published Online, November 2010.

[13] Hou, X., Y. Li and T. Jiang, Fracture Failure Analysis of Ductile Cast Iron Crankshaft in a Vehicle Engine, J. Fail. Anal. and Preven., Published Online, November 2010.

[14] Weckman, G. R., and J. H. Marvel, Forecasting Engine Removal Comparing Neural Network to a Hazard Model, Proceedings of the IIE Annual Conference, May 15-19, Atlanta, Georgia, 2005.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Human Visual Inspection of Automotive Paint Finishes

Edem Tetteh
Virginia State University
etetteh@vsu.edu

Douglas McWilliams
Mississippi State University-Meridian
DMcWilliams@meridian.msstate.edu

Abstract

Customer needs for high-quality goods and the risk of product liability litigation against businesses have made inspection a very important part of manufacturing. The purpose of this research paper was to evaluate the impact of machine-pace and quality criteria on the accuracy of human visual inspection. Subjects inspected two types of automobiles, luxury and inexpensive, under different machine-paced inspection speeds and background density levels in a simulated laboratory experiment.

The study revealed that inspectors spent the least amount of time inspecting paint finishes on inexpensive automobiles with few quality defects under slow paced conditions. However, the inspectors were not accurate in their response selections. In this experiment, inspectors favored speed to the detriment of accuracy and spent approximately the same amount of time searching for the defects under fast pacing as they did with slow pacing.

The results of this study showed that inspections conducted on expensive automobiles cost more than inspections on inexpensive automobiles. Inspections at a fast pace cost the company more than those at a slow pace due to the increased amount of subsequent rework.

1. Introduction

Lean manufacturing has been shown to improve the competitiveness of manufacturing organizations. The concept of lean manufacturing started with the Toyota Production System [10]. Then, the idea was expanded by a research group at the Massachusetts Institute of Technology, starting in the 1980s [17]. The philosophy of lean manufacturing is to use a process of waste reduction, thus producing higher quantities and better quality products with the same or less resources. The goals of lean manufacturing are zero lead time, zero inventory, and zero defects, resulting in higher customer satisfaction [11, 13, 14].

Inspection is often used to attain the goal of zero defects. Inspection is defined as the process of comparing an entire product or part of a product to a related possible standard [1]. Inspections can be done by human inspectors or by automated inspection systems. Automated visual inspection systems have been used extensively for a wide variety of tasks, including blemish detection, complex patterned-surface examination, counting, grading, sorting, sizing, orientation, and posture recognition for robot control [1]. Quality control benefits that arise from the use of automated visual inspections are the improvement of flaw detection, inspection repeatability, reduction of inspection time, electronic capture, the archive of inspection data, and the inspectors' safety [18].

Unfortunately, automated inspection does have some disadvantages and limitations. For instance, the costs to purchase an automated visual inspection system and to train the operators may prove to be too high of an investment for small companies. Moreover, the complexity of some modern, high-tech automated systems may require staffing by highly skilled workers. In addition, the inability to adapt most automated visual inspection systems to new products or targets makes them less attractive. Given the short lifecycle of electronic products and the continuous product mix, this causes the changeover of visual inspection equipment to become very expensive and time consuming. As a result, many companies in the manufacturing industry rely on human inspectors.

Even though there are many wide-ranging studies that involve the extensive use of human

inspectors: electric motors [3], airframe structures [4], and printed circuit boards [8], little research has been conducted to study the effects of human inspection in an automotive assembly facility. Human inspectors are used in such facilities to find and eliminate paint defects on new automobiles. Inspection has been defined as a process or procedure of examining attributes of a part and determining if it does or does not conform to design specifications or quality standards. A relatively common type of inspection task is the type of visual inspection task in which the inspector examines product units one at a time, inspects on one type of defect which may potentially occur at a more or less known location on each unit of product, and decides whether the defect is present or absent [12]. Human inspectors are trained in the art of visually identifying the condition of equipment, tools, products and processes. However, visual inspection is error prone [10], and one of the key determinants of visual inspection performance is visual search. Visual search is influenced by situational awareness, discrimination, memory, and attention. Human visual search behavior tends to be less systematic, which leads to incomplete visual coverage [9]. The advantage of human inspection is the decision-making capabilities of the human inspectors. Humans have the ability to recognize various patterns in quality characteristics and the flexibility to adapt to new situations. This advantage makes humans more effective than automated systems [5]. Because human factors have been shown to be the primary determinants of a visual search task [7], the interest of this study is to identify those human factors that affect the accuracy and reliability of human inspection, particularly pace, density, and quality criteria in automotive paint inspections [15].

The method for determining paint quality is based on the number of allowable paint defects on an automobile not exceeding a maximum number of allowable paint defects. This maximum number, or quality criteria, is selected based on the make and model of the automobile being inspected. For example, in the Nissan assembly plant in Canton, Mississippi, both Nissan and Infinity automobiles are produced on the same assembly line in minimum batch sizes approaching one. However, because Infinity automobiles are expensive automobiles for which customers are willing to purchase at a much higher price, the paint quality for the Infinity automobiles is much higher than for the Nissan automobiles.

Human inspectors have to always be mindful of the automobile type being inspected. If the inspectors apply the wrong quality criteria to inspecting the quality of the paint jobs, the manufacturing costs could be excessive. For instance, if the quality criteria for the Infinity automobiles are applied to the paint jobs of the Nissan automobiles, unnecessary repairs may be ordered, which would increase manufacturing costs. On the other hand, if the paint jobs of the Infinity automobiles are held to the quality standards of the Nissan automobiles, automobiles with undesirable number of paint defects on an Infinity automobile may be passed downstream [16]. The result would be higher repair costs associated with detecting the defects outside of the paint area.

Three research questions are of concern in this study:

1. What impact does pace have on inspection accuracy and speed?
2. What impact does background density have on the conspicuity of blemishes?
3. How does this affect inspection accuracy and speed?
4. What impact does quality criteria have inspection accuracy and speed?

2. Methodology

2.1. Subjects

The participants for this study were thirty undergraduate students enrolled in the College of Technology at Purdue University West Lafayette campus. There were nineteen males and eleven females. They were between 17 and 26 years of age. Nickles et al. have shown that minimal differences exist between inspectors and student subjects on simulated tasks [9]. The

students were not paid for participating in the study.

2.2. Equipment

The experiment was conducted on Pentium IV computers with a Windows XP operating system and a 15-inch monitor with resolution 1280×960 pixels. The subjects viewed the screen from a distance of approximately 500 centimeters. The input devices used to respond to the stimulus material were a Microsoft standard keyboard and a two-button mouse.

2.3. Stimulus material

Simulators were used to predict the performance of existing or planned systems and compared to alternative solutions for particular design problems [2]. The software used to measure the inspection performance was custom designed and used in a previous study by Tetteh et al. [15] to quantify a visual inspection system performance. The interface functions on Microsoft Windows XP operator system. The software was written in Microsoft Visual Basic 6.0, displayed on the screen of the 15-inch monitor. This system included three modules: system setup, system simulation and data collection. The setup module allowed the experimenter to define different inspection parameters: pace and density (Figure 1).

Density represented the percentage of defects. The inspection consisted of examining a number of frames filled with a set of eight characters ("T", "Y", "W", "R", "U", "G", "H', and "P") in the background and searching for a possible random target character, "X."

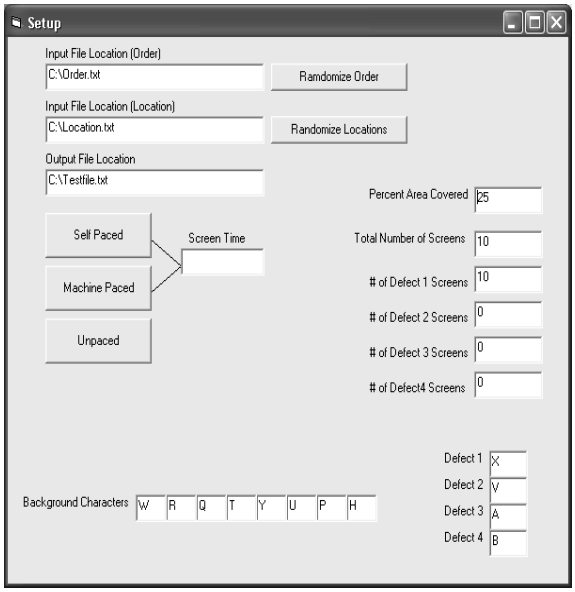


Figure 1: Simulator

The number and locations of the characters were randomly generated, and the random target character "X" represented a paint defect (scratch, blister, pinhole, crater, etc.). The dependent variables were speed and accuracy (hit rate), collected and summarized for each task session.

2.4. Experimental design

A 2×2×2 factorial design was used to analyze the data. The three independent variables were pace, density, and automobile type. Pace had two levels: high pace (7-second display of each frame) and low pace (23-second display of each frame). Two levels were used for density, which were measured by the background density: 75% density (high defect) and 25% density (low defect), illustrated respectively in Figures 2 and 3.

Automobile type also had two levels: expensive and inexpensive (i.e., Nissan Altima and Infinity Q45). A 3-second delay was allowed between the frames. Each subject was exposed to a total of eight possible experimental conditions randomly presented on the computer screen. The study was a between-subject design.

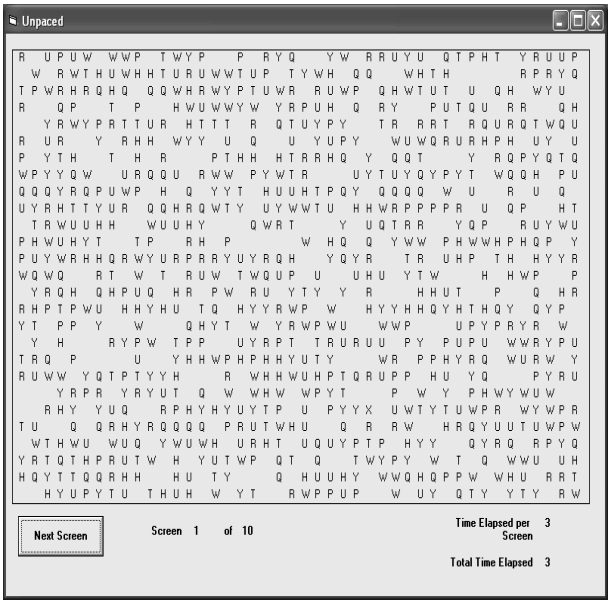


Figure 2. 75% of screen density

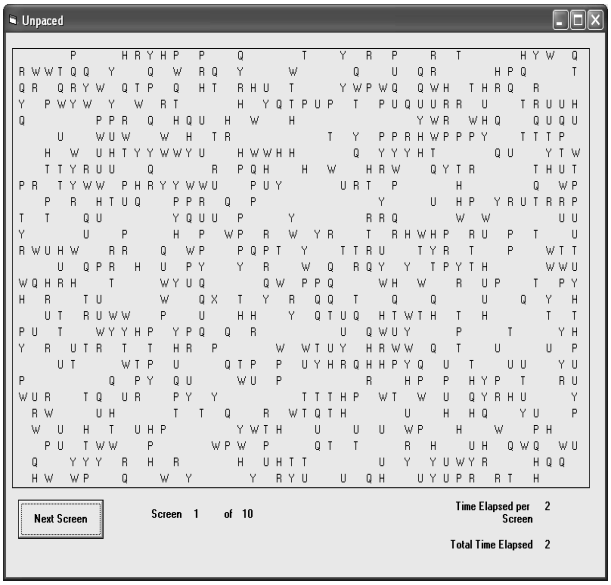


Figure 3. 25 % of screen density

2.5. Procedure

The subjects were first briefed on the nature and objectives of the experiment. Next, they were required to complete a consent form and a demographic questionnaire. After which, the subjects were given an overview of the experiment, which was followed by a program demonstration to familiarize them with the inspection tasks. For each subject, a total of 30 randomly ordered screens with random number of defects were presented. Each screen represented one of the eight possible experimental conditions; the subjects were required to click on each defect "X" when identified or "next" to go to the next screen (or automobile). The subjects were given a signal to indicate the type of automobile each screen represented by displaying a sign with "expensive" or "inexpensive" in each case. They were informed that each screen might or might not have defects: Xs. The average time required by the subjects to complete the visual search tasks was 50 minutes.

3. Results

Data was collected for three performance measures: search time (time taken to locate the targets), accuracy (proportion of defects found), and stopping time (time taken when no target was detected). A mixed factorial ANOVA was used to analyze the three independent variables: pace, density, and automobile type. After confirmation of significant effects or interactions by overall ANOVA, differences among groups were verified by a simple effect ANOVA.

3.1. Search time

A significant interaction effect was found between pace, density, and automobile type ($F(1,376) = 155.35$, $p < 0.0001$); pace and density ($F(1,567) = 57.08$, $p < 0.0001$); and pace and automobile type ($F(1,376) = 25.27$, $p = < 0.0001$). Furthermore, a significant treatment effect was found for density ($F(1,376) = 63.31$, $p < 0.0001$, automobile type ($F(1,376) = 122.38$, $p < 0.0001$) and pace ($F(1,376) = 4.69$, $p = 0.0309$).

Because many interactions were significant a simple main effects test was performed. The result indicated that inspectors spent more time when inspecting expensive automobiles ($F(1,376) = 25.88$, $p < 0.0001$), when inspecting automobiles with high defect($F(1,376) = 93.16$, $p < 0.0001$), and at a slower pace ($F(1,376) = 120.31$, $p < 0.0001$).

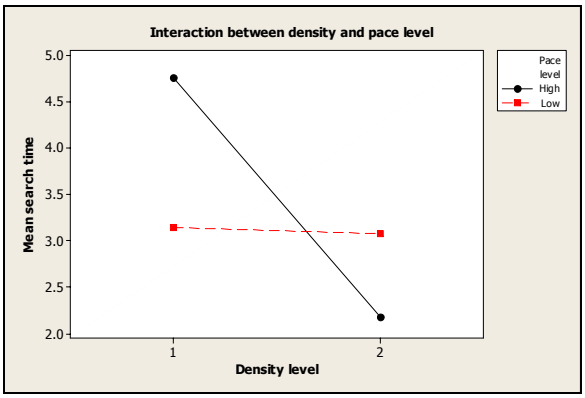


Figure 4. Density x pace

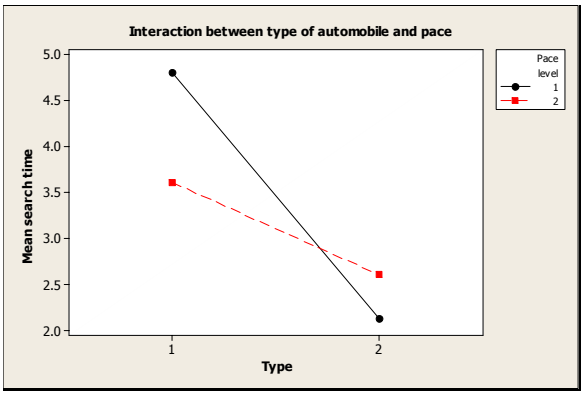


Figure 5. Type of automobile x pace

3.2. Hit rate

A significant interaction effect was found between pace, density, and automobile type ($F(1,376) = 7.84$, $p = 0.0054$), pace and density ($F(1,376) = 15.01$, $p = 0.0001$), density and automobile type ($F(1,376) = 16.44$, $p < 0.0001$).

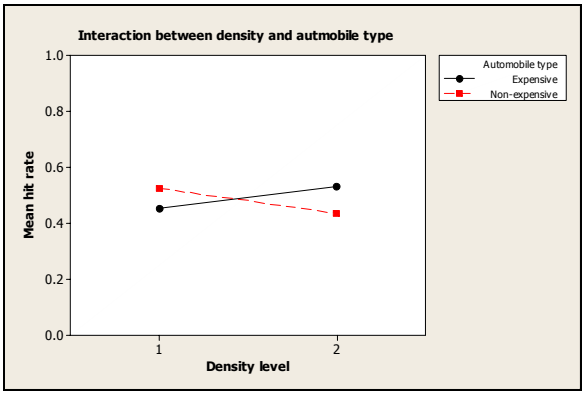


Figure 6. Type of automobile x density

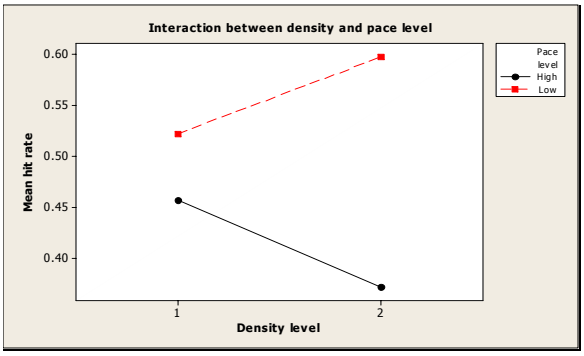


Figure 7. Density x pace

A significant treatment effect was found for pace ($F(1,376) = 49.15$, $p < 0.0001$). A simple main effects test results show that inspectors who performed at a slow pace were more accurate in their decisions ($F(1,376) = 8.49$, $p < 0.0038$).

3.3. Stopping time

A significant interaction effect was found between pace, density, and automobile type ($F(1,376) = 108.19$, $p < 0.0001$), pace and density ($F(1,376) = 18.78$, $p < 0.0001$).

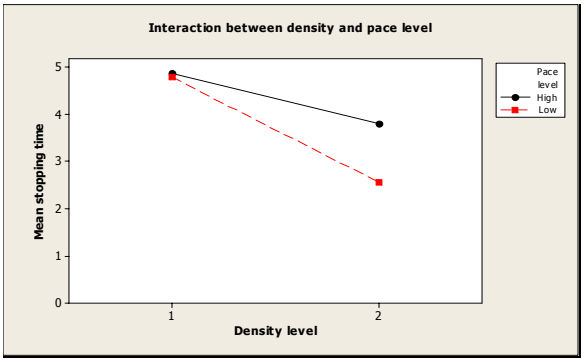


Figure 8. Density x pace

A significant effect was found for automobile type ($F(1,376) = 10.04$, $p = 0.0.0017$), pace ($F(1,376) = 24.77$, $p < 0.0001$) and density ($F(1,376) = 151.11$, $p < 0.0001$). A simple main effects test results show that inspectors who performed under low defect ($F(1,376) = 43.34$, $p < 0.0038$) and at a slow pace ($F(1,376) = 138.21$, $p < 0.0038$) spent less time working on inexpensive automobiles ($F(1,376) = 16.33$, $p < 0.0001$).

3.4. Cost impact

In addition, the economic consequences associated with the inspection were calculated using the following equation (Shanbhag, Sadasiva, & Gramopadhye, 2005):

$$C = \sum\left[\left(ch \times h\right) - \left(cm \times m\right)\right] - kt$$

Where C is the net cost, ch is the cost of detecting the defect, cm is the cost of missing the defect, h is the number of hits, m is the number of misses for a particular defect, and k is the cost-per-unit time, which is \$17 (inspectors' hourly wage) divided by 3600 seconds = \$ 0.0047. Finally, t is the total inspection time. The two treatment conditions for automobile type were assigned with cost values based on inspection outcome, as illustrated in Table 1.

	Detection Revenue	Cost of Failure
Expensive Automobile	\$90	\$60
Less Expensive Automobile	\$34	\$19

Table 1. Cost value structure for automobile types by inspection outcome

The results showed a significant interaction effect between pace, density, and automobile type ($F(1,376) = 36.16$, $p < 0.0001$), density and automobile type ($F(1,376) = 5.63$, $p = 0.0182$), pace and automobile type ($F(1,376) = 19.84$, $p < 0.0001$) and pace and density ($F(1,376) = 43.78$, $p < 0.0001$).

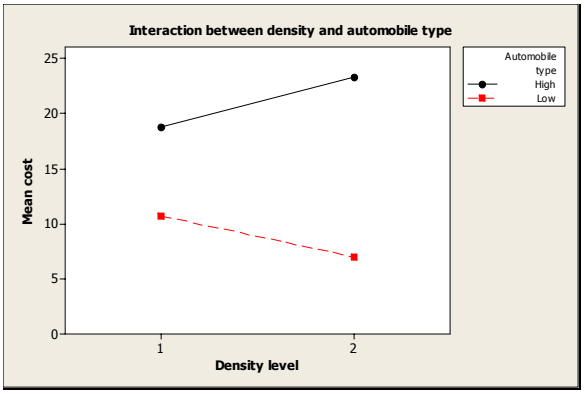


Figure 9. Type of automobile x density

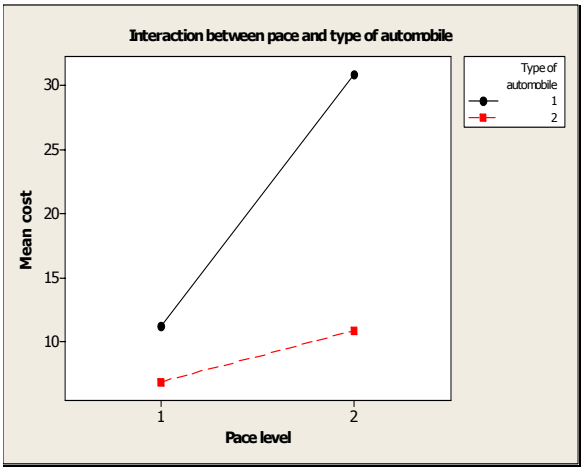


Figure 10. Type of automobile x pace

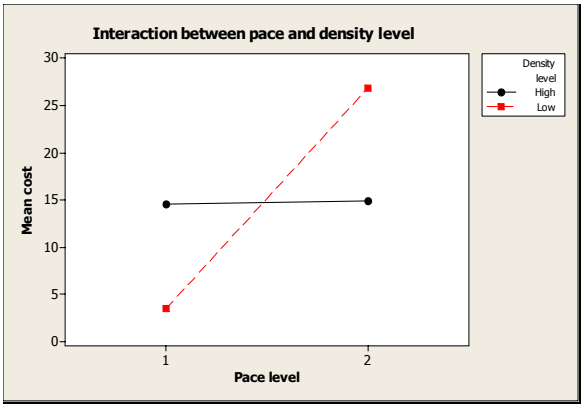


Figure 11. Density x pace

A significant treatment effect was found automobile type ($F(1,376) = 48.60$, $p < 0.0001$), and pace ($F(1,376) = 45.57$, $p < 0.0001$) A simple main effects test results showed that inspection costs were less during slow-pace periods ($F(1,376) = 89.34$, $p < 0.0001$).

4. Discussion

This study was concerned with three research questions. The study aimed to document the possible impact of pace on inspection accuracy and speed. The results showed that there was a significant interaction effect between pace and search time or speed. Inspectors spent more time inspecting expensive automobiles and when inspecting automobiles with high defect at a slower pace. However, the variables could not be isolated and a further study of the

interrelation between pace and speed alone is needed for that purpose. It is generally accepted though that the speed of inspectors is generally dictated by the pace of the production line with inspectors being faster on high pace lines while they have a tendency to take their time on slow pace lines. The study showed that inspectors who performed at a slow pace line were more accurate in their decisions. As such, pace has a direct or positive effect on accuracy regardless of other variables such as density or automobile type. As expected, the accuracy increases as subjects take more time to inspect the screen and they are able to do this more on a slow pace line rather than a high pace line.

The second research question posed at the beginning of this study was the impact of background density on the conspicuity of blemishes. Qualitatively speaking, background density would of course have an indirect effect or negative relationship with the conspicuity of blemishes. Quantitatively this was measured through this study by examining hit rate and density. However other variables were interacting with density and those with pace and automobile type was found to be significant. As such, with more time to inspect and with a greater reason to be more accurate, the conspicuity of blemishes becomes more apparent regardless of density.

In relation to the second research question, the third relates background density with accuracy and speed. In the study it was seen that density negatively affects accuracy and speed with higher density decreasing accuracy and speed and lower density increasing accuracy and speed. The impact of quality criteria in relation to inspection accuracy and speed as the fourth research question was already partly discussed in the second question however it needs to be stressed that the perceived need to be more accurate in inspectors would generally have a trade-off in speed.

Furthermore, the results show that the level of defects affects the change in speed. Inspectors spent approximately the same time searching for the defects under fast pace. Under slow pace with low defect, inspectors had better accuracy. The reason for this is that the fast pace did not give enough time for the inspectors to think through their selections, as the slow pace did. This may also be because students were used instead of real inspectors. Inspections conducted on expensive automobiles cost more than inspections on inexpensive automobiles. Not surprisingly, inspections at a fast pace cost more than those at a slow pace. This can be explained by the fact that fast-paced inspections were not accurate most of the time. Overall, the best performance conditions for inspectors, resulting in fast and accurate inspections that cut costs, is a slow pace, working with inexpensive automobiles that have low defects. This will, however, depend on whether company policy is inclined toward accuracy or speed. The possible explanation is that each one of the condition factors (slow pace, working with inexpensive automobiles that have low defects) is perceived by the inspector as associated with a low stress working environment. As such, the inspectors spent more time, but ultimately were very accurate.

5. Conclusion

This study was able to show that for an automobile lean manufacturing line, pace indirectly affects accuracy while it directly affects speed. Background density has an indirect effect on the conspicuity of blemishes and thus has an indirect effect as well on accuracy and speed. Finally, quality criteria increases accuracy but decreases speed. As with any study, a greater sample size would greatly increase the significance of the results most specifically with regards to accuracy of the results. The limitation of the software use also comes to fore since the main concern in the study is an automobile line where the main characteristic inspected is paint defects. Randomly generated letters on the screen does not do this justice. Since computers were already used for this study a better test would be to show colors or pictures that have variations in the solid color indicating a defect. Simulating the actual inspection and production line even through a computer would greatly increase the accuracy of the study.

Future investigation will take into account some important factors omitted from this study, such as the use of skilled inspectors and the addition of other treatments, such as search

strategy, contrast level, noise level, or lighting level. These aspects are important to fully comprehend and train visual inspectors to find a central point between speed and accuracy and to also come up with means to successfully inspect expensive automobiles with high defects at low cost.

6. Reference

[1] Batchelor, B.G., Hill, D.A., & Hodgson, H., Automated visual inspection. New York: Elsevier Science Publications, 1985.

[2] Benedettini, O. & Tjahjono, B., "Towards an improved tool to facilitate simulation modeling of complex manufacturing systems", International Journal of Advanced Manufacturing Technology. 2008, 43(1/2), pp. 191-199

[3] Chaney, F.B. & Teel, K.S., "Improving inspector performance through training and visual aids", Journal of Applied Psychology, 1976, 51, pp. 311-315.

[4] Drury, C., Prabhu, P., & Gramopadhye, A.K., "Task analysis of aircraft inspection activities: Methods and findings", Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA, 1990, pp. 1181-1185.

[5] Drury, C.G., & Sinclair, M.A., "Human and machine performance in an inspection task", Human Factors, 1983, 25, pp. 391–399.

[6] Garrett, Sandra K. Melloy, Brian J. & Gramopadhye, A.K., "The effects of per-lot and per-item pacing on inspection performance", International Journal of Industrial Ergonomics, 2001.

[7] Gramopadhye, A.K., Drury, C.G., & Prabhu, P., "Training strategies for visual inspection", Human Factors and Ergonomics in Manufacturing, 1997, 7(3), pp. 171–196.

[8] Jiang, X., Bingham, J., Master, R., Gramopadhye, A.K., & Melloy, B., "A visual inspection simulator for hybrid environments" International Journal of Industrial Engineering: Theory, Applications and Practice, 2002, 9(2), pp. 123-132.

[9] Nickles, G.M., Melloy B.J., & Gramopadhye, A K., "A comparison of three levels of training designed to promote systematic search behavior in visual inspection", International Journal of Industrial Ergonomics, 2003, 32 (5), pp. 331-339.

[10] Ohno, T., Toyota Production System. New York: Productivity Press, 1998.

[11] Sanchez, A.M. & Perez, M.P., "Lean indicators and manufacturing strategies", International Journal of Operations & Production Management, 2001, 21(11), pp. 1433- 1451.

[12] Smith, L.A. & Barany, J.W., "An elementary model of human performance on paced visual inspection tasks", IIE Transactions, 1970, 2(4), pp. 298-308.

[13] Soriano-Meier, H. & Forrester, P.L., "A model for evaluating the degree of leanness of manufacturing firms", Integrated Manufacturing Systems, 2002, 13(2), pp. 104-109.

[14] Tapping, D., Luyster, T., & Shuker, T. Value Stream Management. New York: Productivity Press, 2002.

[15] Tetteh E., Jiang X., Mcbride M., Mountjoy D. & Seong Y., "Evaluation of job-aiding tool in inspection systems", Human Factors and Ergonomics in Manufacturing, 2008, 18(1) pp. 30-48

[16] Whitfield, K., Smyrna maxes out: Automotive Design & Production, Gardner Publications, Inc., 2002.

[17] Womack, J.P., Jones, D.T., & Roos, D. The Machine that Changed the World, New York: Harper Perennial, 1990.

[18] U.S. Department of Transportation, Automation inspection aircraft, National Technical Information Service, Springfield, VA, DOT/FAA/AR-97169, 1998.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Intelligent Framework for Software Analysis, Reuse and Fabrication

Paulus Wahjudi
Marshall University
wahjudi@marshall.edu

Abstract

Software re-engineering is necessary for several reasons. These include fixing defects (corrective reengineering), modifying the software to address weaknesses or to mitigate potential malfunctions (preventive reengineering), and extending the software to accommodate changes in its external environment (adaptive reengineering). The objective of the paper is to introduce a framework that provides a common repository structure to represent information about existing software assets and their operating environments. The goal is to facilitate information exchange related to the transformation of existing software assets. The framework will provide the ability to document existing systems, discover reusable components in existing software, and support transformations to new systems. The framework will also provide information about existing software artifacts to be exchanged among different systems. This will allow for a flexible software solution that addresses the different levels of expertise within the development team.

1. Introduction

Software engineering involves more than the development and deployment of software. Once it has been deployed, software must be maintained. Software maintenance is defined as the process of changing the software after it has been deployed [10]. Software maintenance is an important part of the software lifecycle that requires constant support. Upkeep can cost more than the original price of the software with older software tending to be much more costly to maintain [6]. The primary cost of maintenance tends to arise from the many customization requests and add-ons performed on the basic software package. The various additions often introduce new bugs and errors that generate various unwanted side effects. These side effects can be either positive or negative, although in most cases there is a high probability that the side effect will be negative.

Maintenance involves a number of activities including reverse engineering (creating design or other documentation artifacts from the source code) and reengineering. Reengineering involves using forward engineering techniques to add new functionality or develop a completely new system based on an original system's specifications [11]. Reengineering arises from a variety of needs including: (1) corrective reengineering needs i.e., the need to fix defects, (2) perfective reengineering needs i.e., the need to change the software to enhance its overall performance, (3) perfective reengineering needs i.e., the need to change the software to prevent potential malfunctions, and (4) adaptive reengineering needs i.e., the need to adapt the software to changing hardware or software environments [10]. Reengineering decisions are driven by the need to improve the flexibility of existing systems and reduce operational expenses through workflow optimizations. In cases where formal documentation does not exist, or where existing documentation has been rendered obsolete, software re-documentation is an important part of the reengineering process.

In this paper, the Intelligent Software Reengineering Framework (ISRF) will be proposed to assist in the process of software reengineering. It will be shown that it will reduce the time needed to perform modification and additions to existing software. Section 2 of the paper will describe the framework components and the role of the each component in the reengineering process. A preliminary

implementation and analysis will are presented in section 3 with conclusions and proposals for future work on the ISRT in section 4.

2. Intelligent Software Reengineering Framework

The behavior of software systems is inherently difficult to decompose and isolate while maintaining clear understanding of the application's logic. One of the first research activities in identifying crosscutting concerns in source code was performed by Hannemann and Kiczales [6]. Their Aspect Mining Tool (AMT) supports both text and type-based analysis. Moreover, it can be extended to other types of analysis, such as signature based searches. Since each of these analyses has benefits and drawbacks, AMT has been set up as a "multi-modal analysis tool", permitting the combined use of different techniques. Each analysis works as a query and results in a series of matched lines that are visualized in the system's source code. The "Aspect browser" tool [5] is based on the assumption that aspects, defined as secondary design decisions, have a signature (textual-pattern, lexical token). This signature is a textual regular expression that will help in aspect identification. Similar to the text based analyzers of AMT, the success of the tool strongly depends on naming conventions followed by the authors of the analyzed code. The location of the code that implements the aspects s is represented graphically by views that are based on the Seesoft concept [4]. Aspects can also be identified by using dynamic program analysis techniques. This approach was proposed by [2], where certain classes of aspects and their dynamic trace patterns are defined. The mining process searches for these patterns in program traces that were captured during execution. The main classes considered are *outside-aspects*, that is the call of method 'a' is always followed by the call of method 'b', and *inside-aspects*, that is the call of method 'b' is always inside a call of method 'a'. They also distinguish subclasses such as *outside-aspects* that are *before* or *after* a specific method call, and *inside-aspects* that are *first* or *last-in* a specific method call. The research of Rashid and Loughran [7] addresses the question of how aspects found in an existing system should be represented. The approaches considered are the *storing of the aspect as an object* (binary/character) together with the object's description (meta-data) that can be queried, the *mapping of the aspect anatomy to a database model* that will allow the storing of aspect's properties, and the *hybrid approach*, a combination of the previous two. The mapping of the aspects to a relational database model is presented in more detail in [8]. Aspect mining is used by Zhang and Jacobsen [12] to conduct quantitative analysis of the tangling and scattering phenomenon in existing legacy implementations of middleware architectures (CORBA). New aspects specific to the chosen platform and defined as abstractions that crosscut the major architectural components considered by the authors are reported. These are the dynamic programming model and portable interceptors. The Aspect Oriented Programming (AOP) based re-factorization of a number of aspects identified was performed, and the quantified results aim to demonstrate a reduction of the complexity and improvement of performances compared to the original implementation. The tool used in the analysis, AMTEX4, was built on top of the AMT tool described earlier.

The Intelligent Software Reengineering Framework (ISRF) consists of several key action components:

- Identify Files, Folders, Features and Functionalities (I4F).
- Assessment of System Connectivity, Reciprocity and Validity (A3Y).
- Framework Catalog Compilation (F2C).
- Software Taxonomy Mapper (STM).

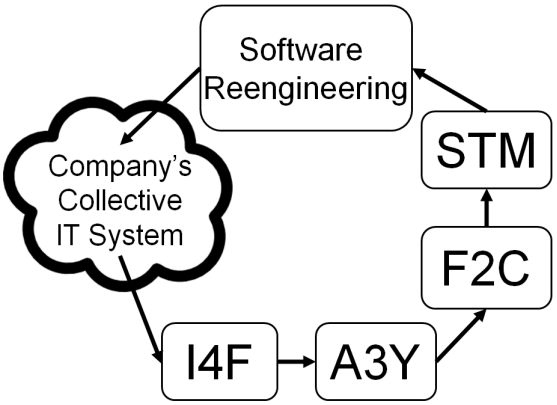


Figure 1. Intelligent Software Reengineering Framework

The framework utilizes a step by step approach to continuously refine and produce an accurate understanding of an existing system. It addresses the complexity among various modules that work together to support a company's daily activity. Figure 1 illustrates the process of creating an ISRF to support the software reengineering process. A company's collective Information Technology (IT) system is represented as a big cloud of programs containing various elements capable of performing various tasks. The first step is to identify the various components of the system by using I4F. Once completed, A3Y will perform analysis of I4F result and certify that the results are accurate and valid. F2C will then store the results in appropriate categories. The latter allows for quick retrieval of components and reasoning necessary for the reengineering process. STM will then retrieve data pertinent to the reengineering and provide a representation of the various components that will likely be affected by the reengineering process.

2.1 Identify Files, Folders, and Features and Functionalities

All the key features in software and aspect mining discussed are incorporated into ISRF. The first step of our framework works by generalizing the analysis. This includes the identification of the files and folders associated with the current system. I4F contains two steps as shown in figure 2. The first is the identification of various components' locations and their respective type. The Paths locator mechanism coupled with the programs detector will peruse every possible location for elements of the company's IT system. This process is critical to identify the system components and to compensate for the lack of documentation of any software that is currently in use.

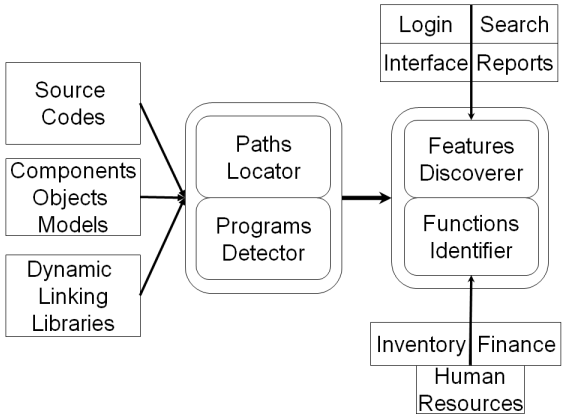


Figure 2. Identify Files, Folders, Features & Functionalities

The second step is to determine the function and features of the components that were identified in the first step. A function is defined as a set of identifiable units in the company that utilize the system such as inventory, financial, and human resources. Each unit is strongly connected to the others. Similarly, features are defined as specific set of common tasks that are performed by the system. There are currently four basic main features: login, search, interface and report. The task of discovering features and identifying functionalities is a combination of an automated system and a human expert. The features discoverer searches for specific keywords and phrases within a program and provides the list to be authenticated by a human expert. For example, a login procedure is associated with required username and password

while a search process requires user to enter a set of keywords or topics. The function identifier focuses primarily on how the system interacts with the data. Each unit often deals with a specific dataset and does not modify data outside its scope.

2.2 Assessment of System Connectivity, Reciprocity and Validity

Once I4F is done, the connection and dependence between various components in the software must be identified and checked. Each file defined by I4F is treated as an element. These elements are interconnected and a weighted value is given to the connections. Highly dependent elements will create stronger bonds. A3Y explore every element and identifies the elements that are connected and the weighted values associated with them. The process of identifying and calculating the weighted connection is done based on the features and functionalities determined by I4F. In an attempt to ensure every dependency is considered, A3Y will traverse every possible flow of execution within the system utilizing both white box and black box testing methods. A3Y will examine different code independent paths. It will then derive a set of execution paths that guarantee the following: (1) each line of code in a module is executed at least once, (2) all logical decisions are executed for both *true* and *false* case, (3) all iterative segments are executed within their specified boundaries. However, over analysis has a risk of over complicating the connections between elements. Thus A3Y devises a weighted value in each connection. This threshold value will adjust the number of point to point connections available on the nodes providing a more concise representation.

2.3 Framework Catalog Compilation

Cataloging the internal structure of software makes it easier to understand and cheaper to modify without changing its observable behavior. In F2C, the elements discovered in I4F and A3Y are systematically organized into catalogs. The procedure results in a relational database. Each file is logged based on its function, feature and identified associations.

There are three main data tables in F2C: (1) the SOURCES table containing the information regarding the source files and it location, (2) the FEATURES table that list all the possible features that can be discovered by I4F, (3) the FUNCTIONS table list all of the identified function the company. To support the connection between the main tables, a set of connecting tables is defined to represent connectivity and reciprocity. Each table is populated with the data generated by I4F with respect to the represented value. The connections are referred through A3Y.

2.4 Software Taxonomy Mapper

The Software Taxonomy Mapper mines F2C for information and creates a representation of the current system. It does so by representing each identified components from I4F as a node. The connection along with the weighted value defined through A3Y is represented as a node connection. The knowledge inference step in most software reengineering, which “invents” the higher levels of abstraction is underdeveloped: existing aspect mining work concentrates on how to distill concern-related data from the sources (signatures, lexical pattern matching), and how to present such data via, for example, visualization. These steps may support the human engineer in aspect mining, but are unable to come up with any aspect themselves. Additionally, most knowledge inference techniques are often overwhelmed due to the massive amount of information that they attempt to process and result in inadequate analysis. To overcome this problem, a knowledge inference engine is built into STM allowing for a generalized representation with all the key

features but without the overhead of existing tools that process raw data.

Each element is also represented as a node. Then a node graph can be constructed with nodes and their respective connections. A set of pre-defined anchor nodes identified by a human analyst will mark the unique area for each function. Each node with a strong association to particular functionality will be positioned in an area around the anchor nodes while maintaining its connectivity with other nodes. This approach is used to create a manageable flocking pattern of files with similar functionalities. Using the graph, we then identify specific groupings or flocking patterns that consistently emerge upon mapping - such as database connection, report generation, user inputs, and processing libraries. STM utilizes anchor nodes that are nodes that have been confirmed to contain a specific set of functionality and features that are highly connective to other nodes. With the aid of the anchor nodes the overall map is grouped to produce unique and identifiable patterns. The identified patterns are then examined to determine a common identifiable type of software, for example: an inventory control system, storefront interface, payroll, etc.

3. Implementation and Analysis

The ISRF methods presented above were applied to an existing project. The project was based on build based on various computer science courses projects with four major functionalities – input, database, libraries, and reports. The analysis was done by senior-level Computer Science students under the author's direction. It was developed using Java and was applied to a project also written in Java. First, the I4F examined a designated home folder and parse any Java files in that folder. Then subsequent subfolders were investigated to produce a tree of the file structure. Second, the I4F Features Discoverer and Functions Identifier parsed every Java file found. Both components utilized a user defined Regular Expression to parse the Java file. Each valid parsing indicated the existence of a feature and/or a function within that file. The file was then tagged for review. Once I4F had gone through the entire repository, a report was generated to be reviewed by an expert analyst. One of the main problems found in this testing step was the amount of false positives and redundant identification found by I4F. This resulted in a large number of tagged files. We managed to reduce the problem by refining the Regular Expression. We foresee that in the future, I4F will need to incorporate some form of computational intelligence to identify the various files.

Once the tags were confirmed by the human analyst, we performed a white box testing on the project as part of A3Y. Similar to I4F, the main obstacle encountered in this step was the labor intensive demand to perform the tests. We intentionally avoided utilizing any commercial white box testing software as the testing results generated by such software do not meet our needs. However, this problem will need to be addressed as the size of the projects grow and tax the limits of manual analysis.

F2C then cataloged every tagged file along with the functions and features associated with it. We utilized a MySQL database connected through JDBC with three primary tables: tblFiles, tblFunctionalities and tblFeatures. The Files table contained details of each Java file such as the filename and file path along with a unique fileID. Similarly, the Functionalities and Features table contained a list of every feature and functionality tagged in the system and a unique ID number.

The STM component queries the database and generates a graphical representation of the entire repository on a Java canvas. Each file was represented as a node and the dependency between files was represented as a connecting line. A representation of a node graph generated

by STM is shown in figure 3. Based on the tag associated with the file, the nodes are placed in a specific predetermined location. Depending on the number of files in the repository and the level of interconnectivity, the drawing process requires a considerable amount of time. The process is not computing intensive but relies heavily on communication with the database.

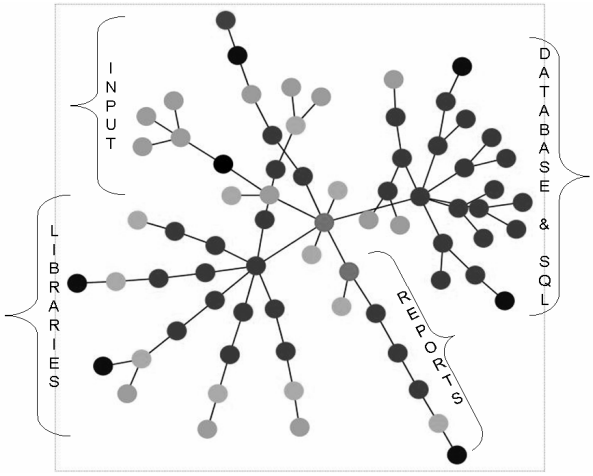


Figure 3. Software Taxonomy Map

The final results of the validation experiments are shown in Figure 3. The graph shows each file and which of the four functionalities it was most closely associated with. Using this diagram would allow developers and managers to identify which files are most closely related to the new task.

4. Conclusion & Future Work

ISRF is an attempt to extend the highly generalized software exploration method to the overly complicated knowledge inference approach. We do so by limiting the amount of software exploration necessary to identify key components that allows for sufficient knowledge inference to be made based on the pattern generated. The utilization of ISRF in software reengineering will reduce the time necessary to reverse engineer existing components, software landmarks and various interactions between components providing support for future evolution of the system. Once ISRF is in place, the I4F and A3Y will be able to continuously monitor changes within the system while F2C and STM will be able to provide accurate representations of the current state of the IT system. In time, ISRF generates a continuous cycle of analysis and refinement that will increase the efficiency and efficacy of the company IT system. Preliminary testing and analysis of framework yielded various problems that need to be addressed. Although ISRF is currently in the testing and development stage, the framework proves to be effective in aiding software developers to analyze existing software system. These problems primarily affected performance and efficiency of the framework while the overall results are promising.

To further refine the framework, we will need to address the deficiencies in the various components of ISRF especially in the I4F's ability to better identify the various features and functions without yielding too many false positives. Although a human analyst will always be needed to verify the findings, devising a more accurate and intelligent screener will significantly reduce the analyst workload and increase efficiency. Testing of the ISRF will also need to be expanded to incorporate complex systems and possibly other programming languages.

In the future, we are also planning to expand ISRF to support software development. The information generated by ISRF can provide developers the necessary knowledge about existing system and the necessary components related to the new software. This knowledge includes predicting the possible components to be affected by the addition of the new software whether it is due to functional redundancy, supported features, or other aspects.

5. References

[1] Agarwal, R., Sarangi, A., and Das, S. "Reengineering of database intensive application", SIGSOFT Softw. Eng. Notes 28, 3 (May. 2003), 1-1.

[2] Breu, S. and Krinke, J. “Aspect mining using dynamic analysis.” In Workshop on Software-Reengineering, Bad Honnef, 2003.

[3] Deshpande, Y., Chandrarathna, A., and Ginige, A. “Web site auditing: first step towards re-engineering”, In Proceedings of the 14th international Conference on Software Engineering and Knowledge Engineering (Ischia,Italy, July 15 - 19, 2002). SEKE '02, vol. 27. ACM, New York, NY, 731-737.

[4]Eick, S.G, Steffen, J.L and Sumner, E.E. “Seesoft – tool for visualizing line oriented software.” IEEE Transactions on Software Engineering, 18(11):957–968, November 1992.

[5] Griswold, W.G, Kato, Y. and Yuan, J.J. “Aspect browser: Tool support for managing dispersed aspects.” Technical Report CS1999-0640, 3, 2000.

[6] Hannemann, J. and Kiczales, G. “Overcoming the prevalent decomposition of legacy code.” In Proc. Workshop on Advanced Separation of Concerns. IEEE, 2001.

[7] Loughran, N. and Rashid, A. Mining aspects. In Workshop on Early Aspects: Aspect-Oriented Requirements Engineering and Architecture Design, Enschede, The Netherlands, 2002.

[8] Loughran, N. and Rashid, A. “Relational database support for aspect-oriented programming.” In Proceedings of NetObjectDays Conference, number 2591 in LNCS, pages 233–247, 2002.

[9] Pressman, R. “Software Engineering A Practitioner’s Approach”, 7th Ed., Chapter 29, McGraw-Hill, 2010.
[10] Sommerville, I. “Software Engineering”, 8th Ed, Addison-Wesley, 2007.

[11] Van den Brand, M. G., Klint, P., and Verhoef, C. “Reverse engineering and system renovation—an annotated bibliography”, *SIGSOFT Softw. Eng. Notes* 22, 1, Jan. 1997, 57-68.

[12] Zhang C. and Jacobsen. H-A. “Quantifying aspects in middleware platforms.” In Proc. 2nd Int. Conf. on Aspect-Oriented Software Development (AOSD), pages 130–139. ACM Press, March 2003.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

A Hybrid Parallel Kangaroo & Simulated Annealing Algorithm for Multi-Objective Flow Shop Scheduling

Mustafa Kerim Yılmaz, Alpaslan Fığlalı, Ümit Terzi
University of Kocaeli, Turkey
m_kerim@hotmail.com; figlalia@kocaeli.edu.tr; umitterzi@kocaeli.edu.tr

Mesut Yavuz
Shenandoah University
myavuz@su.edu

Orhan Engin
University of Selçuk, Turkey
orhanengin@yahoo.com

Abstract

In this paper a hybrid parallel kangaroo-simulated annealing algorithm is proposed to solve a flow shop scheduling problem with three minimization objectives: the makespan, total completion time and total machine idle time. The paper also develops a new weighting approach to combine the objectives to obtain more realistic and balanced weights. For the solution of the emerging multi-objective flow shop scheduling problem, a novel hybrid parallel kangaroo - simulated annealing algorithm is proposed. The effectiveness of the proposed algorithm is demonstrated thru a computational study based on benchmark problems from the existing literature. Computational results prove that the hybrid parallel kangaroo-simulated annealing algorithm is promising.

1. Introduction

The flow shop scheduling problem is one of the most general job scheduling problems. In a flow shop environment, one is concerned with optimizing the flow of n jobs thru m machines. Each job requires m operations, one on each machine and to be performed in the same sequence, i.e., operation j of job i (i = 1, 2, .., n) is performed on machine j (j = 1, 2, …, m). The processing time of operation j of job i, t_{ij}, is finite and known in advance. Furthermore, jobs are independent and simultaneously available at time zero, processing times are not time or sequence dependent, and a machine can process at most one job at a time.

Flow shops are utilized in various industries where a number of products need to go through a common sequence of operations. A flow shop environment is more complicated than a single machine due to the number of operations involved. On the other hand, it is less complicated than a job shop owing to the existence of a common sequence. Assembly lines can be considered simplified flow shops where stations are balanced so that each job requires a unit time in each station, i.e., machine. Flow shops are widely seen in hi-tech production systems, such as electronics manufacturing, where operations cannot be efficiently distributed to stations in a balanced way. For more detailed accounts of flow shop scheduling, including various versions of the problem, see [1-2].

A special case of flow shop scheduling problem is the permutation flow shop scheduling problem. In permutation flow shops the same job

sequence is followed on all machines. In other words, a production sequence is established for the first machine and all the other (m-1) machines process the jobs in a first-come, first-served basis, thereby preserving the sequence set at the first machine.

In a majority of the papers studying the flow shop scheduling problem, the objective is the minimization of makespan. However, there exist several other objectives such as total flow time and total machine idle time that are considered as important as makespan in various flow shop systems. This paper is concerned with three minimization objectives: makespan, total completion time and total machine idle time in flow shop scheduling.

The (permutation) flow shop scheduling problem is well-known to be NP-Hard even with only the makespan objective and three machines [3]. Since no one has been able to find a polynomial time exact algorithm to solve an NP-Hard problem, it is quite unlikely that a fast-running exact optimization algorithm exists for the flow shop scheduling problem. Therefore, (meta-)heuristic approaches have become increasingly popular for the solution of permutation flow shop scheduling problems in recent years [4]. Following this line of research, this paper proposes a new hybrid meta-heuristic solution method for multi-objective permutation flow shop scheduling.

The remainder of this study is organized as follows: In Section 2, permutation flow shop scheduling objectives are discussed. The parallel kangaroo algorithm is presented in Section 3 and the hybrid parallel kangaroo and simulated annealing algorithm is proposed in Section 4. A computational study and its results are given in Section 5. Section 6 concludes the paper with a summary and discussion of future research directions.

2. Permutation flow shop scheduling objectives

Permutation flow shop scheduling problems studied in the literature include

- Makespan
- Total (Weighted) Completion Time
- Total (Weighted) Flow Time
- (Total or Maximum) Earliness and/or Tardiness
- (Weighted) Number of Tardy Jobs
- Total (Weighted) machine idle time

The makespan denotes the completion of the last job in the sequence is arguably the most commonly studied of these objectives. Given a finite number of jobs, the makespan tells us when the flow shop system completes all assigned jobs and becomes idle again. In real-life, new jobs arrive often before all the sequenced jobs are completed, thereby creating a need to frequently re-optimize the production sequence. Total completion time objective emphasizes completing as many of the jobs as early as possible so that the flow shop system gets a natural relaxation towards the end. This is more beneficial in many real-life systems than minimizing the makespan.

Total flow time is very similar to the total completion time in purpose. In fact, if all jobs are available at time 0, the two objectives are identical. Tardiness is important when a due date is given for each job. Traditionally, early completion has been welcomed and only late completion has been penalized. With the spread of the just-in-time philosophy, earliness has become increasingly important, giving rise to earliness-tardiness objectives in scheduling.

All these objectives focus on jobs. In contrast, total machine idle time objective emphasizes machines and measures their idle time as a source of waste in a flow shop system.

For all of these objectives, except for the makespan, different mathematical functions can be formulated to incorporate weights (i.e.,

importance or priority levels) of jobs and/or machines.

This paper is concerned with the makespan, total flow time and total machine idle time objectives. The notation of the multi-objective flow shop scheduling problem considered in this paper is given as follows.

- $t_{i,j}$ = processing time of job i on machine j
- $c_{i,j}$ = completion time of job i on machine j
- C_i = completion of job i ($C_i = c_{i,m}$)
- $J(s,k)$ = Job in the k^{th} position of sequence s

Objectives:

- The makespan value *(M)*
  $M = C_{J(s,n)}$ (1)
- Total completion time (*C)*
  $C = \Sigma_i C_i$ (2)
- Total machine idle time (*I*)
  $I = \sum_j \sum_k \left(C_{J(s,k),j} - t_{J(s,k),j} - C_{J(s,k-1),j} \right)$ (3)

There are a few meta-heuristic solution techniques developed for multi-objective flow shop scheduling problems in the literature. Arroyo and Armentano [5] proposed a partial enumeration heuristic for flow shop scheduling problems with (i) makespan and maximum tardiness and (ii) makespan and total tardiness objectives. Then, Arroyo and Armentano [6] developed a genetic local search algorithm for the multi-objective flow shop scheduling problems again with (i) makespan and maximum tardiness and (ii) makespan and total tardiness objectives. Qian et al. [7] proposed a differential evolution based hybrid algorithm for multi-objective permutation flow shop scheduling problem with makespan and maximum tardiness objectives. Geiger [8] presented a novel study of the structure of Pareto sets with application to multi-objective flow shop scheduling problem. Later, Geiger [9] presented a study of the problem structure of multi-objective permutation flow shop scheduling problems and investigated the effectiveness of local search neighborhoods within an evolutionary search framework. The objective functions studied in [8] and [9] were the makespan, total completion time, average completion time, total tardiness, maximum tardiness, number of tardy jobs, and total machine idle time. Shi et al. [10] proposed a hybrid escalating evolutionary algorithm for solving flow shop scheduling problems with (i) makespan and total tardiness, and (ii) makespan, average flow time and average tardiness objectives. Li and Wang [11] presented a hybrid quantum inspired genetic algorithm for multi-objective flow shop scheduling problem with makespan and maximum tardiness objectives. Yagmahan and Yenisey [12] proposed an ant colony algorithm for the flow shop scheduling problem with makespan, total flow time and total idle time objectives. Engin et al. [13] developed a parallel kangaroo algorithm for the same multi-objective flow shop scheduling problem. Later Yagmahan and Yenisey [14] proposed a multi-objective ant colony system algorithm for flow shop with the objectives of makespan and total flow time.

In all these studies, multiple objective functions are either combined by using pre-determined weights or used jointly in search of a Pareto-optimal front. When weights are used to combine multiple objective functions into only one, a common practice is to assign equal weights.

A closer examination of the objectives reveals that *C* takes on values much larger than *M* and *I*. Thus, equal weight allocation distorts the combined objective as *C* dominates it. The hybrid meta-heuristic proposed in this paper calculates a weight for each objective function specifically for the instance at hand to guide the search process.

To the best of our knowledge, this paper is the first study using instance-specific objective function weights to guide the search process, and distance from an ideal solution to measure solution quality.

3. Parallel kangaroo algorithm

The kangaroo algorithm was first introduced by Pollard in 1978 [15]. The kangaroo algorithm is an approximation technique based on stochastic descent, and has been successfully used for solving optimization problems. Serbencu et al. [16] proposed a hybrid meta-heuristic for solving the single machine scheduling problem by combining an ant colony system and the kangaroo algorithm. Minzu and Beldiman [17] proposed a hybrid genetic algorithm – kangaroo method to solve (i) single machine scheduling and (ii) tasks-to-workstations assignment problems. Öztürk et al. [18] developed a parallel kangaroo algorithm for solving job shop scheduling problems. Engin et al. [13] proposed a parallel kangaroo algorithm for solving multi-objective flow shop scheduling problem with the makespan, total flow time and total idle time objectives. Yılmaz et al. [19] proposed a parallel kangaroo algorithm for solving no-wait flow shop scheduling problems with the makespan as the single objective.

The kangaroo method is an iterative procedure which minimizes an objective function $f(x)$. A current solution x of the considered problem is replaced by a better one, located in its neighborhood $N(x)$, using a random selection. The algorithm tries k times to improve the current solution, where k is a parameter of the algorithm. If a new improvement is no longer possible, a "*jump*" procedure is performed, in order to escape from the attraction of a local minimum. Improvement of the current solution is not obligatory and a different neighborhood definition $N'(x)$ can be used for random jumping.

In the parallel kangaroo algorithm, a tame and a wild kangaroo cooperatively search for the optimal solution. The cooperation takes place at the end of each iteration as the wild kangaroo cues the best solution it has found to the tame kangaroo. Thru this parallel search, a larger number of solutions can be effectively explored.

The local search method embedded in the kangaroo algorithm is the well-known *hill climbing* method, in which, one updates the current solution if and only if a better neighbor solution is found. This method suffers from a major deficiency, i.e., getting trapped in local optima. In the next section, a hybrid method overcoming this deficiency is proposed.

4. A hybrid parallel kangaroo & simulated annealing algorithm

In this paper, a surrogate objective function, O, is used to guide the search process to solve the multi-objective flow shop scheduling problem under consideration.

For each instance of the problem, a large number of solutions are created at random and means of the three objective functions are computed. Then, the combined objective function is calculated with weights inversely proportional to means;

$$O = \frac{M}{a_1} + \frac{C}{a_2} + \frac{I}{a_3} \tag{4}$$

where

$a_1 = \overline{M}$,

$a_2 = \overline{C}$, and

$a_3 = \overline{I}$.

For example, let the means of the three objective functions be 100, 500 and 200, respectively. Then, the combined objective function value of a solution is calculated by

$$O = \frac{M}{100} + \frac{C}{500} + \frac{I}{200}.$$

Clearly, an average solution would have a combined objective value of 3, and better solutions would have smaller objective values. In this approach, the contributions of different objectives are balanced and relative improvements are sought instead of unit improvements.

The proposed hybrid parallel kangaroo-simulated annealing algorithm benefits from a few principles that are common with some other

meta-heuristics: (i) multi-start local search, (ii) multiple different neighborhoods, and (iii) information sharing.

The wild kangaroo makes a jump at the end of each iteration, thereby giving each iteration a fresh start. This gives the method a multi-start local search flavor.

The wild and tame kangaroos use two separate neighborhood functions. The wild kangaroo searches for solutions that are less similar to the current solution, whereas the tame kangaroo searches for closer solutions to the current. This is a common feature with contemporary meta-heuristics such as variable neighborhood search and iterated local search.

At the end of each iteration the wild kangaroo shares the best solution it found during the iteration with the tame kangaroo. If that solution is better than the current solution of the tame kangaroo, the tame kangaroo makes a move. Clearly, this is a form of information sharing. Collective intelligence such as possessed in ant colony and particle swarm optimization heavily relies on information sharing.

In the described structure of the parallel kangaroo algorithm, the wild kangaroo represents *exploration* and the tame kangaroo represents *exploitation*. Performance of meta-heuristic algorithms depends on achieving a good balance between exploring a broad set of neighborhoods and exploiting the neighborhoods of the best solutions found in the process.

In this study, an insert neighborhood is used. A position in the sequence is chosen at random. The job in the chosen position is removed and then inserted back into the sequence at another position. The distance between the first and last positions is the kangaroo's step size. Figure 1 illustrates this process for a kangaroo with a step size of four.

Simulated annealing is a trajectory-based meta-heuristic method [20]. It aims to escape local optima by probabilistically allowing non-improving moves. The probability of accepting a non-improving move depends on the energy level, which decreases with time.

The simulated annealing method has been successfully implemented on numerous optimization problems, see [21] for a recent survey. In this paper, simulated annealing is used as the local search mechanism in the parallel kangaroo algorithm. Both kangaroos search their neighborhoods using simulated annealing with different starting energy levels and cooling functions.

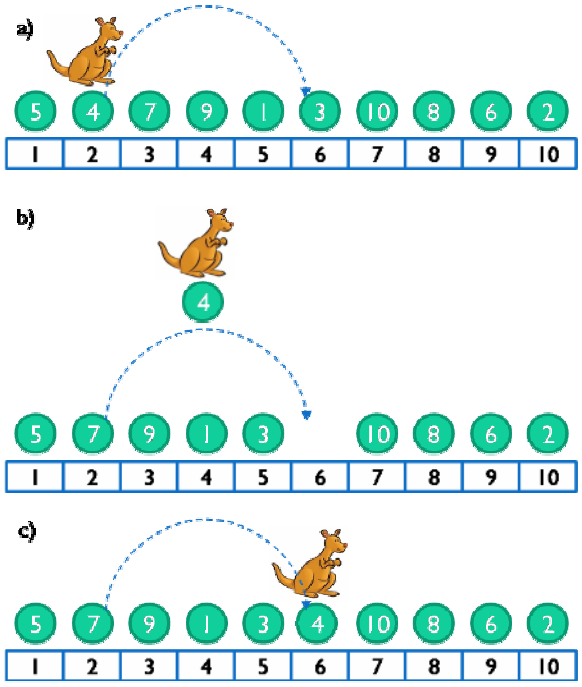


Figure 1. Random jumping of a kangaroo

A pseudo-code of the proposed hybrid parallel kangaroo-simulated annealing method is given as follows.

1. Randomly create a starting solution for each kangaroo and compute its objective function value.
2. Randomly choose a solution in the wild kangaroo's neighborhood.
3. If the new solution is better, make the wild kangaroo jump to it. Otherwise, make the

jump probabilistically as in simulated annealing.

4. If the wild kangaroo has not jumped *Wild Kangaroo # Jumps* times, return to step 2.
5. Randomly choose a solution in the tame kangaroo's neighborhood.
6. If the new solution is better, make the tame kangaroo jump to it. Otherwise, make the jump probabilistically as in simulated annealing.
7. If the tame kangaroo has not jumped *Tame Kangaroo # Jumps* times, return to step 5.
8. If the wild kangaroo's solution is better than that of the tame kangaroo's, pass wild kangaroo's solution to the tame kangaroo.
9. Randomly create a new solution for the wild kangaroo and compute its objective function value.
10. If the number of iterations has reached *Max # Iterations*, then terminate; otherwise, return to step 2.

5. Computational study

Twenty one test problems are chosen from the existing literature [22]. The problem sizes vary from 20 jobs x 5 machines to 75 jobs x 20 machines. The performance of the proposed parallel kangaroo and simulated annealing algorithm (PK-SA) is compared with Engin et al.'s [10] parallel kangaroo (PK-I) and weighted parallel kangaroo algorithm (PK-II). Note that PK-II uses instance-specific weights for objective functions but is not hybridized with simulated annealing. In all three algorithms the objectives are makespan, total flow time and total machine idle time and a_1, a_2, and a_3 are the weights for each objective respectively. In PK-I equal relative weights for the objectives are chosen in combined objective value ($a_1 = a_2 = a_3 = 0.333$). In PK-II and PK-SA the objectives are combined in weighted form with the new approach that normalizes the magnitudes and the weights are given in Table 1.

The proposed hybrid parallel kangaroo & simulated annealing algorithm is implemented with the following parameter values:

i) parallel kangaroo parameter values
   Wild Kangaroo Step Size = 4
   Wild Kangaroo # Jumps = 70
   Tame Kangaroo Step Size = 2
   Tame Kangaroo # Jumps = 20
   Max # Iterations = 200

ii) simulated algorithm parameter values
   Initial temperature = 100
   Boltzmann's constant = 0.95

Table 1. Weights for PK-II and PK-SA

Problem	a_1	a_2	a_3
ReC01	0.392	0.092	0.516
ReC03	0.416	0.100	0.484
ReC05	0.369	0.085	0.546
ReC07	0.615	0.077	0.309
ReC09	0.616	0.077	0.307
ReC11	0.607	0.075	0.318
ReC13	0.737	0.065	0.197
ReC15	0.743	0.067	0.190
ReC17	0.747	0.068	0.185
ReC19	0.602	0.074	0.324
ReC21	0.619	0.078	0.303
ReC23	0.596	0.073	0.331
ReC25	0.722	0.062	0.216
ReC27	0.711	0.060	0.230
ReC29	0.725	0.063	0.212
ReC31	0.567	0.067	0.366
ReC33	0.584	0.070	0.346
ReC35	0.558	0.066	0.376
ReC37	0.755	0.047	0.198
ReC39	0.741	0.045	0.215
ReC41	0.737	0.044	0.218

The algorithm is coded in Borland Delphi and run on a PC with an Intel Core i5 750 processor with 2.66 GHz and 8 GB memory. All methods are evaluated based on the deviation of their solution from an ideal solution. The ideal solution is a hypothetical solution that has the objective function values M^*, C^* and I^* that are obtained by solving the test instance with a

single objective at a time. The deviation of a solution from the ideal then is calculated by:

$$\Delta = \sqrt{\left(M - M^*\right)^2 + \left(C - C^*\right)^2 + \left(I - I^*\right)^2} \qquad (5)$$

Table 2 shows the deviations from ideals for the 21 test problems.

Table 2. Deviations from ideal solutions

Problem	*nxm*	PK-I	PK-II	PK-SA
ReC01		193.79	189.59	167.10
ReC03	20x5	219.07	108.13	84.49
ReC05		197.70	189.57	183.80
ReC07		414.06	354.42	285.04
ReC09	20x10	1,169.46	738.42	700.24
ReC11		830.94	669.28	414.45
ReC13		1,352.11	1,273.92	1,029.56
ReC15	20x15	2,797.78	2,243.57	1,632.49
ReC17		883.54	809.10	594.94
ReC19		693.63	631.54	561.38
ReC21	30x10	975.92	974.95	819.24
ReC23		1,048.30	1,008.83	485.26
ReC25		2,299.09	1,255.40	678.35
ReC27	30x15	1,705.48	1,110.21	962.58
ReC29		1,549.71	1,478.36	1,147.61
ReC31		729.24	690.80	520.76
ReC33	50x10	1,119.00	844.24	810.62
ReC35		962.93	955.58	955.31
ReC37		4,466.44	4,103.07	3,744.97
ReC39	75x20	7,477.64	4,591.99	2,789.62
ReC41		6,026.83	5,949.32	2,564.12
Avg.		**1,767.29**	**1,436.68**	**1,006.28**

The relative deviation from the best known solution for any algorithm (Alg) is calculated, as a percentage, as follows:

$$\%(\text{Alg}) = \frac{\Delta(\text{Alg}) - \min(PKI, PKII, PKSA)}{\min(PKI, PKII, PKSA)} x100$$

The percent relative deviations for all algorithms are given in Table 3.

As seen in Table 3, over all twenty one problems, PK-I has yielded 69.42%, PK-II 35.46% and the proposed PK-SA 0.00% average relative deviation. That is, for all test instances, the proposed hybrid PK-SA algorithm has found the best solution among the three methods tested. Between PK-I and PK-II, we notice that PK-II performs better on all 21 test instances.

These results prove that both of the ideas proposed in the paper are effective: (i) assigning balanced weights to multiple objective functions in guiding the search process leads to better solutions than assigning equal weight, and (ii) hybridization of the parallel kangaroo algorithm with simulated annealing outperforms the simple parallel kangaroo algorithm.

Table 3. Percent relative deviations

Problem	*nxm*	PK-I	PK-II	PK-SA
ReC01		15.97	13.46	0.00
ReC03	20x5	159.28	27.98	0.00
ReC05		7.56	3.14	0.00
ReC07		45.27	24.34	0.00
ReC09	20x10	67.01	5.45	0.00
ReC11		100.49	61.49	0.00
ReC13		31.33	23.74	0.00
ReC15	20x15	71.38	37.43	0.00
ReC17		48.51	36.00	0.00
ReC19		23.56	12.50	0.00
ReC21	30x10	19.13	19.01	0.00
ReC23		116.03	107.90	0.00
ReC25		238.93	85.07	0.00
ReC27	30x15	77.18	15.34	0.00
ReC29		35.04	28.82	0.00
ReC31		40.03	32.65	0.00
ReC33	50x10	38.04	4.15	0.00
ReC35		0.80	0.03	0.00
ReC37		19.27	9.56	0.00
ReC39	75x20	168.05	64.61	0.00
ReC41		135.04	132.02	0.00
Avg.		**69.42**	**35.46**	**0.00**

6. Conclusion

In this paper a hybrid parallel kangaroo-simulated annealing algorithm is proposed to solve a multi-objective flow shop scheduling problem. 21 of Reeves's [22] benchmark problems are used as the test-bed to test the methods proposed in the paper. Computational results prove that the proposed hybrid parallel kangaroo-simulated annealing algorithm has found the best solution for all twenty one instances in the literature.

In a future study, a larger base of test instances may be solved to generalize the results obtained here. Also, in the search process, Pareto fronts obtained by different methods may be used to compare the methods instead of only the final solutions.

7. References

[1] M. Pinedo, "Scheduling: Theory, algorithms and systems", 3rd edition, Springer, 2008.

[2] J.N.D. Gupta, and E.F. Stafford, Jr., "Flowshop scheduling research after five decades", European Journal of Operational Research, 169(3), 2006, pp. 699-711.

[3] M.R. Garey, D.S. Johnson, and R. Sethi, "The complexity of flowshop and jobshop scheduling", Mathematics of Operations Research, 1, 1976, pp. 117-129.

[4] R. Ruiz, and C. Maroto, "A comprehensive review and evaluation of permutation flowshop heuristics", European Journal of Operational Research, 165(2), 2005, pp. 479-494.

[5] J.E.C. Arroyo, and V.A. Armentano, "A partial enumeration heuristic for multi-objective flow shop scheduling problems", Journal of the Operational Research Society, 2004, 55(9), pp. 1000-1007.

[6] J.E.C. Arroyo, and V.A. Armentano, "Genetic local search for multi-objective flowshop scheduling problems", European Journal of Operational Research, 2005, 167, pp. 717- 738.

[7] B. Qian,. L. Wang, D. X. Huang, and X. Wang, "Multi-objective flow shop scheduling using differential evolution", Intelligent Computing in Signal Processing and Pattern Recognation, Book series, Springer, Verlag, Berlin, 2006, pp. 1125-1136.

[8] M.J. Geiger, "On the distribution of Pareto optimal solutions in alternative space- the investigation of multi objective permutation flow shop scheduling problems", Technological and Economic Development of Economy, XII, 2006, 1, pp. 23-29.

[9] M.J. Geiger, "On operators and search space topology in multi-objective flow shop scheduling", European Journal of Operational Research, 181, 2007, pp. 195- 206.

[10] R.Shi, Y. Zhou, and H. Zhou, "A hybrid escalating evolutionary algorithm for multi-objective flow shop scheduling", Third international conference on natural computation, 2007, ICNC, IEEE computer society.

[11] B.B. Li, and L. Wang, "A hybrid quantum inspired genetic algorithm for multiobjective flow shop scheduling", IEEE Transactions on systems, Man and cybernetics-Part B, Cybernetics, 2007, 37, pp. 576- 591.

[12] B. Yagmahan, and M.M. Yenisey, "Ant colony optimization for multi-objective flow shop scheduling problem", Computers & Industrial Engineering, Elsevier, 2008, 54(3), pp. 411- 420.

[13] O. Engin, M.K. Yılmaz, A. Fığlalı, and N. Fığlalı, "A parallel kangaroo algorithm for multi-objective flow-shop scheduling", 6th International Symposium on Intelligent and Manufacturing Systems, 2008, Sakarya, Turkey.

[14] B. Yagmahan, and M.M. Yenisey, "A multi-objective ant colony system algorithm for flow shop scheduling problem", Expert systems with applications, Elsevier, 37, 2010, pp. 1361- 1368.

[15] J.M. Pollard, "Monte Carlo methods for index computation (mod *p*)", Mathematics of Computation, 32, 1978, pp. 918-924.

[16] A. Serbencu, V. Minzu, A. Serbencu, "An ant colony system based metaheuristic for solving single machine scheduling problem", The Annals of "DUNAREA DE JOS" university of GALATI FACCICLE III, 2007.

[17] V. Minzu, L Beldiman., "Some aspects concerning the implementation of a parallel hybrid metaheuristic", Engineering Applicants of Artificial Intelligence, 20(7), 2007.

[18] A. Öztürk, M.K. Yılmaz, O. Engin, "A parallel kangaroo algorithm for solving job shop scheduling problem", 28. Operations Research and Industrial Engineering National Congress, 2008, Galatasaray University, İstanbul, Turkey.

[19] M.K. Yılmaz, A. H. Kökçam, V. Duvarcı, A. Fığlalı, M. Ayöz, and O. Engin, "A parallel kangaroo algorithm for solving no-wait flow shop", 29. Operations Research and Industrial Engineering National Congress, June 2009, Bilkent University, Ankara, Turkey.

[20] S. Kirtpatrick, C.D. Gelatt, Jr. and M.P. Vecchi, "Optimization by simulated annealing", Science, 220, 1983.

[21] E. Aarts, J. Korst, and W. Michiels, "Simulated Annealing", in Search Methodologies, E. Burke and G. Kendall (Eds.), Springer, 2005, pp. 187-210.

[22] C.R. Reeves, "A Genetic algorithm for flow shop sequencing", Computers and Operations Research, 22(1), 1995, pp. 5- 13.

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Fault Detection for Smart Grid Reliability Enhancement

Logan Porter, Joshua McCracken, Joshua Kirksey, Dylan Kwan, Wei Zhan, and Joe Morgan
Texas A&M University
zhan@entc.tamu.edu

Abstract

A student capstone project team is sponsored by Oncor, a power distribution company, to develop a smart sensor for circuit breakers to detect several different types of faults. The sensor is required to be nonintrusive, low cost, work in harsh environments with a high level of electromagnetic interference, and communicate with both a remote central controller and a technician using a laptop through digital communication protocols RS-232 and RS-485. The smart sensor greatly enhances the reliability of smart grid power transmission systems. This paper discusses the detail of the design and the fabrication of the smart sensor used to monitor circuit breakers.

1. Introduction

Smart grid technology is changing the way electrical power is generated, distributed, and used. The enhanced capability of the power grid due to the use of digital communication and control can greatly improve efficiency, reduce the probability of power outage, and allow for early detection and prognosis of faults of the power grid system. Federal, state and local governments are investing tens of billions of dollars in the nation's infrastructure for smart grid. Texas A&M University anticipated the new trend in smart grid technology and is collaborating with the power industry to enhance the curriculum and technology research capability in this area.

A key component of the smart grid system is smart sensors that can predict or detect potential problems and send warnings to the central controller or engineers using digital communication. This paper disc usses the design of a smart sensor that has the fault detection and prediction capability for circuit breakers. This is a part of a capstone design project sponsored by Oncor, an electric distribution and transmission company that serves three million homes and businesses in Texas.

In the power transmission world, circuit breakers are used to prevent damage to equipment in the power grid [1-3]. A fault current is typically caused by either a short to ground or by two power lines contacting each other. Both of these faults cause the current on the line to rise to unsafe levels that can damage other power transmission equipment. When a fault occurs, a circuit breaker is used to open the circuit on which the fault occurred so that customers' and the company's properties are not damaged. Many states have mandatory reliability reporting requirement and some have penalties for excessively poor reliability [4].

Fig. 1 shows the basic operation of a circuit breaker in a substation. When a circuit breaker does not open the circuit that just experienced a fault condition, breaker failure has occurred. When a circuit breaker failure occurs, it is important to quickly diagnose the problem with the breaker so that repairs can be made and power can be restored to the customer. Currently, there are limited devices or methods that can diagnose the problem with the breaker, let the power transmission company know immediately that breaker failure has occurred, and accurately describe the problem. More importantly, prediction of circuit breaker fault

allows the utility companies to order and place the circuit break before it fails.

The smart sensor CBM09 designed by the capstone project team T.I.E.S. [5] is a smart monitoring electronic device that is installed in the cabinet of a circuit breaker. The CBM09 module uses sensors to monitor the key functions of the circuit breaker and conduct analysis of the data collected by these sensors. The module can reliably determine if a problem has occurred or is about to occur with the operation of the circuit breaker. In these cases, the CBM09 will raise an alarm to be sent to the control room, indicate the problem with a series of LEDs mounted on its case, and provide detailed information to a technician connecting to the module with a laptop.

2. Design process

The team followed a system engineering approach [6, 7] to design the smart sensor. The design process started with collecting the Voice of Customers. Discussion with the sponsor and other utility companies were conducted to establish the appropriate design requirements. The team then developed a conceptual design. Subsystem and component level requirements were derived from the system level requirements. A test plan was designed to validate the requirements from component level all the way to system level [8]. Project management tools such as Critical Path Method [9] were used during the entire project.

2.1. System requirements

The system requirements are divided into three categories: monitoring, user interface, and others.

Monitoring

The smart sensor shall monitor the following

- Status of the trip coil:

  The trip coil's current shall be monitored. The device shall be checking for an open or short circuit across the trip coil.

- Trip signature:

  The device shall monitor the trip signature for abnormal trip, which is when the trip signature exceeds the upper or lower limits. It shall also monitor for a slow trip, which is when the time required to trip exceeds the maximum time allowed.

- Re-strike:

  The device shall monitor for re-strike, which is when the current continues to flow through the breaker after the contacts of the circuit breaker have been separated.

- DC ripple:

  The DC voltage supplied to the device shall be monitored for an AC component on the signal.

User Interface

- DC voltage:

  The CBM09 shall allow the user to select between 125V and 48V DC for the level of voltage being supplied to the CBM09. This option allows the CBM09 to be used with different kinds of circuit breakers because substations use either 125V or 48V to power the equipment which includes the circuit breakers.

- 3 or 5 cycle trip time:

  The CBM09 shall allow the user to select between three and five cycle trip times. The selection determines the maximum time allowed for a slow trip to occur since a five cycle trip time takes longer than a three cycle trip time.

- Sensitivity:

  The CBM09 shall allow the user to select between a low and high percentage tolerance, which determines how much the trip signature that was just recorded can vary from the reference trip signature at any given time.

- Clear reference waveform:

The CBM09 shall allow the user to clear the reference waveform. This will help when configuring the CBM09 in different circuit breakers.

- Reset:

  The reset shall clear all LED indicator lights. Reset shall not restart the program.

- DIP switches:

  The CBM09 must provide a switch so that the user can make a selection for each of the parameters above.

- PC software for accessing and viewing trip signatures:

  The device shall allow the user to download the captured data to a laptop and view it graphically.

- RS-485 and RS-232 communication ports:

  There shall be two communication ports to transmit data from the device. The RS-485 port [10] is used to send alarms to the control room which can relay it back to central control. The RS-232 port allows the technician to connect to the device with a laptop and download operational data on the circuit breaker.

Others

- Size:

  The device shall not exceed 10" x 8" x 3" based on space constraints inside the cabinets of circuit breakers.

- Operational temperature range:

  The device shall function properly from -20°C to 85°C.

- Temperature range for storage:

  The device shall not be damaged for environmental temperatures from -30°C to 100°C.

- Cost:

  The per-device cost for 100 devices shall be less than $1,000.

2.2. Conceptual design

The conceptual block diagram in Fig. 2 displays a pictorial overview of how the CBM09 module works. The functional diagram of the smart sensor module CBM09 is illustrated in Fig. 3.

The CBM09 is placed inside the circuit breaker and uses sensors to measure the current across the trip coil of the circuit breaker and the phases of the breaker. The current measurements are processed by a microprocessor to determine if any fault has occurred during a trip. The information about the fault is relayed to the substation control room by RS-485 digital communication protocol.

The CBM09 also allows for user interfacing using RS-232 and user input switches. The RS-232 port allows a user to interface a laptop with the device to collect a data log of trip signatures. The switches allow the user to configure the CBM09 for different types of circuit breakers in a substation. This allows the CBM09 to work universally with current as well as with older circuit breakers.

2.3. Test plan

After the establishment of requirements, a test plan was created for validation of the requirements at components, subsystems, and system levels. As the design precedes, component level and subsystem level tests will be conducted to validate the design requirements. A system level functional test will then be performed in the field at a substation site. Table 1 illustrates how each test is related to specific requirements. The test plan includes the purpose of each test, test equipment to be used, test procedures, and pass/fail criterion.

2.4. Schematics and board layout

The module can be powered from either a 125 V or a 48V DC source. This power is stepped down using a resistor and a Zener diode to approximately 25 V. Voltage regulators with

output of 15 V and 3.3 V are used to create constant voltages required by a Hall Effect sensor and the dsPIC304011 microcontroller and RS-232 and RS-485 IC chips. Non-intrusive current transformers were selected for sensing the three phase currents through the circuit breaker. The signals from the sensors go through signal conditioning to turn the AC output voltage to a readable DC voltage by the PIC microcontroller. A non-intrusive Hall Effect sensor detects the current through the trip coil when a trip has occurred. This signal is passed through a filter and amplifier to be read on an analog PIC input channel. The microcontroller analyzes the signals and communicates the status of the circuit breaker with the control room via RS-485 or a technician's laptop via RS-232 communication protocols. A slow blown fuse is used between the DC power supply of the circuit breaker and the CBM09 module for protection of the circuit breaker from potential faults, such as short circuit, inside CBM09.

Schematics and board layout were created for CBM09 using Multisim/Ultiboard [11]. Fig. 4 shows the current sensing circuitry. This includes three phase detector circuits and a signal conditioning circuit for the Hall Effect sensor. The current transformers are connected to a load resistor and passed through a half wave rectifier to eliminate negative voltages. The maximum voltage is capped at 3.3V to keep the PIC input voltage from being too large. The output from the Hall Effect sensor is filtered, scaled down between 0.6 to 3 volts and fed through a buffer before entering the PIC analog input channel.

Fig. 5 depicts the power circuit for the CBM09. It is configured for 125 or 48 VDC. The 125 V configuration uses a 470 ohm and three 47 ohm power resistors and a 25 V Zener diode to supply appropriate current and voltage to the 15 V and 3.3 V regulators. Fig. 6 shows the layout of the dsPIC30f4011. The DIP switches are pulled low when not connected and use debouncing capacitors to prevent noise. The various connections from LEDs, sensors, and ICs are also shown. Fig. 7 shows the communication setup. A RS-232 and RS-485 IC are used and connected to a header that leads to a DB9 connection on the enclosure. The RS-485 protocol used is a 2 wire half duplex Modbus.

The schematic design was thoroughly simulated and analyzed in Multisim. Fig. 5 illustrates a simulation set up for the DC power conversion and voltage regulation. The worst case condition in terms of temperature and tolerance for each component was considered to ensure that the CBM09 module will work under all conditions specified in the system requirements.

Fig. 8 shows the board layout in Ultiboard. In the design, the power, analog, digital, and communication components were separated and placed so that interference from one component to another would be minimized. The printed circuit board (PCB) was built and went through bench testing and initial testing in a breaker as illustrated in Fig. 9.

2.5. Software flowchart

The overall software flowchart is illustrated in Fig. 10. The software was developed in C using MPLAB IDE [12]. The software first initializes variables and I/O pins being used. It then loads a reference waveform, if available, and enters the closed state loop. In this loop, the CBM09 is monitoring DC ripple, trip coil continuity, and any communication requests. If a trip occurs, the device enters the open state loop. In this loop, if there is no reference waveform, the CBM09 stores the waveform as a reference. If there is a reference, the current captured waveform is analyzed to the reference. The CBM09 then waits for communication requests and for the breaker to be closed before going back to the closed loop state.

2.6. Cost analysis

Cost analysis for the CBM09 can be described as a financial breakdown of labor hours, workspace, hardware parts, software design suites, packaging, test equipment, and printed media.

In designing the project, various hardware options were considered. The largest cost of the

hardware was the power components. The prices of DC converter modules were examined for their ease of use, but the cost was determined by the T.I.E.S. team to be too high. The team designed a low cost version of the power converter using surface mount components. Table 2 shows a hardware cost comparison between the original design and the improved design. The cost was reduced from $128 to $14.50 for the power converter. This amounts to a reduction of 89% in cost for the power converter.

Software suites were chosen that would cost the least for the project team. These included using academic versions of National Instruments Circuit Design Suite [11], MPLAB C30 compiler [12], and other programming and testing software readily available for the team.

3. Results

The CBM09 design went through several revisions based on the results from component level, subsystem level, and system level tests specified in the test plan. The first problem identified was the power rating of the resistors in the power conversion circuit. The problem was solved with a combination of proper selection of the resistors and the addition of a heat sink. The second problem identified during field test was the noise in the current sensor signal. The problem was solved with digital filter design with proper cut-off frequency. Fig. 11 shows the data acquisition of the CBM09 module using LabVIEW. The test data correctly shows the normal closing of a circuit breaker. The capability of detecting and predicting a fault condition of a circuit breaker is crucial to the reliability of the power distribution system. This kind of intelligent device will be a key part of smart grid systems.

4. Conclusion and discussion

A low cost smart sensor for monitoring the health of circuit breakers was designed and fabricated by a capstone project team at Texas A&M University. The team followed the system engineering approach for the development of the smart sensor. Performance, cost, non-intrusiveness of the sensor, and robustness against temperature were taken into consideration in the design process. The smart sensor provides flexibility in power supply voltage, user interface, and data communication. A prototype was fabricated and extensively tested to verify that all design requirements were met. It was successfully tested in the field (Fig. 12). Viability of commercial product development based on this prototype is being evaluated. Through this project, the students gained valuable experience in designing a product, managing the project, and testing and fabricating the prototype. Exposure to product development process can greatly reduce the gap between what the students learn and what they will face in the real world [13]. This project also enhanced the collaboration between the Electronics Engineering Technology program at Texas A&M University and local power industry.

Acknowledgement

The authors would like to thank Oncor for their sponsorship of the capstone project, all the technical support, and test equipment provided during the two-semester capstone project.

References

[1] R. E. Brown, Electric Power Distribution Reliability. New York: Marcel Dekker, 2002.

[2] J. J. Burke, Power Distribution Engineering: Fundamentals and Applications. New York: Marcel Dekker, 1994.

[3] Electrical Distribution System Protection, Cooper Power Systems, Milwaukee, WI, 1990.

[4] C. A. Warren and M. J. Adams, “Reliability on the regulatory horizon,” in Proc. IEEE Transm. Dist. Conf. Expo., Atlanta, GA, Oct. 2001.

[5] J. Kirsey, L. Porter, J. McCracken, D. Kwan, “T.I.E.S. capstone design final report”, Texas A&M University, May, 2010.

[6] N. B. Reilly, Successful Systems Engineering, New York: Chapman & Hall, 1993.

[7] A. Kossiakoff and W. N. Sweet, Systems Engineering Principles and Practice. Wiley-Interscience, Dec, 2002.

[8] J. Kirsey, L. Porter, J. McCracken, D. Kwan, "T.I.E.S. test plan", Texas A&M University Capstone Design report, Feb., 2010.

[9] H. Kerzner, Project Management, A System Approach to Planning, Scheduling, and Controlling. 9th ed., Wiley, 2005.

[10] Modbus application protocol specification, V1.1b, Modbus-IDA, Dec. 28, 2006.

[11] Getting start with NI circuit design suite, National Instruments, 2007.

[12] Microchip Technology Inc., MPLAB IDE User's Guide, 2005.

[13] W. Zhan, B. Zoghi, and R. Fink, "The Benefit of Early and Frequent Exposure to Product Design Process", Journal of Engineering Technology. Spring 2009, pp. 34-43.

Figures and tables

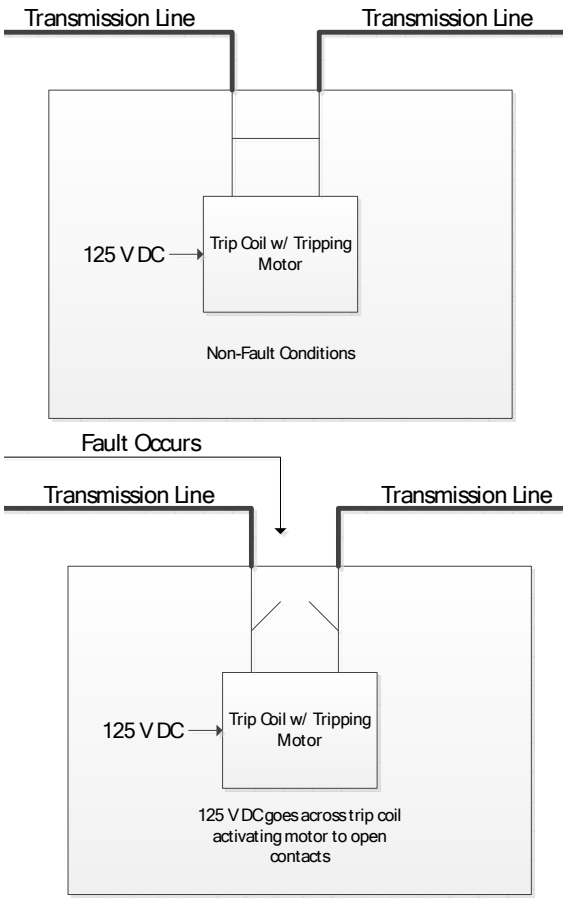


Fig. 1. Basic circuit breaker operation

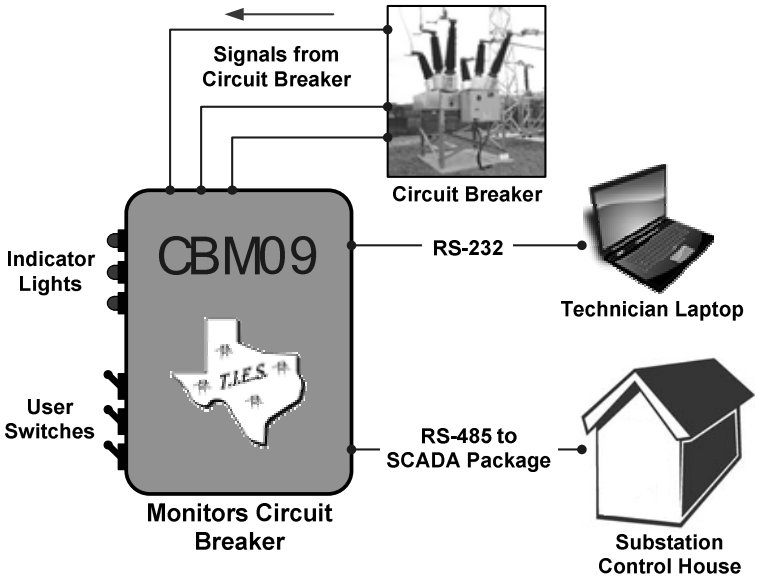


Fig. 2. Conceptual Block Diagram

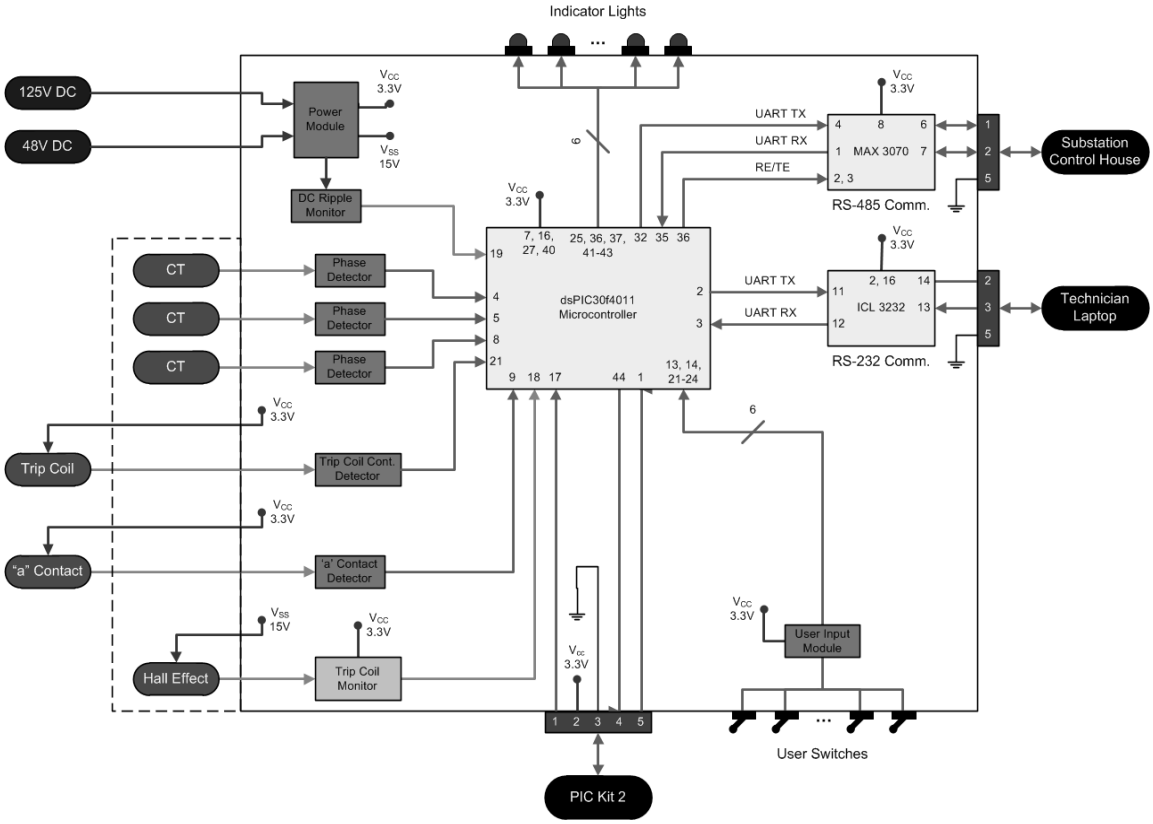


Fig. 3. Functional Diagram

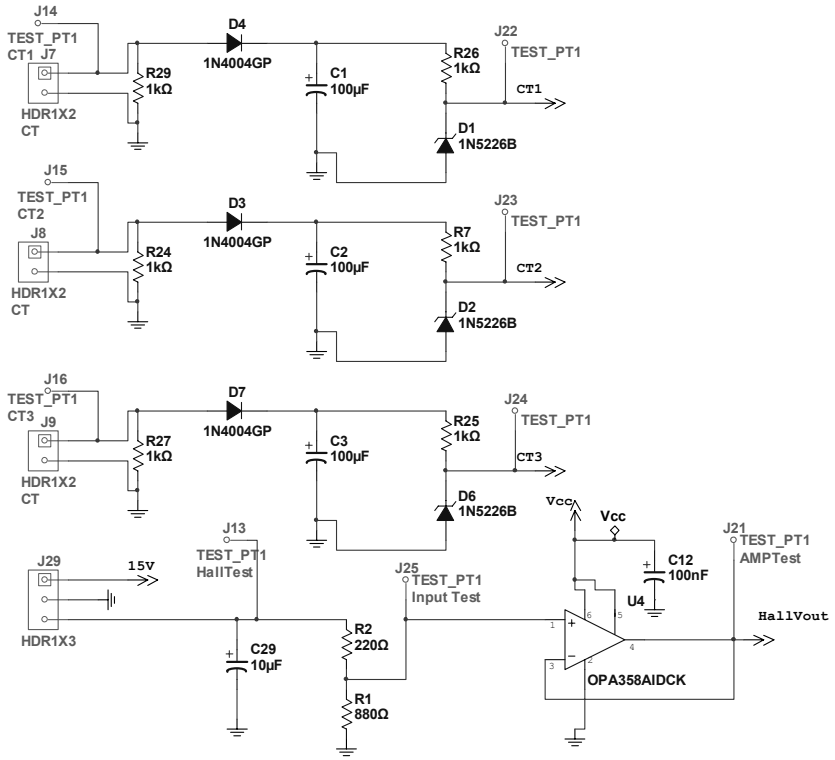


Fig. 4. Current sensing circuitry

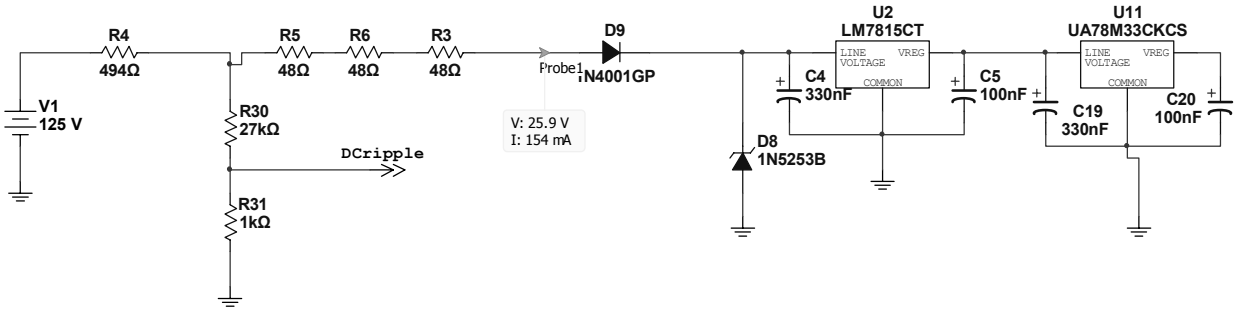


Fig. 5. Voltage regulation circuitry

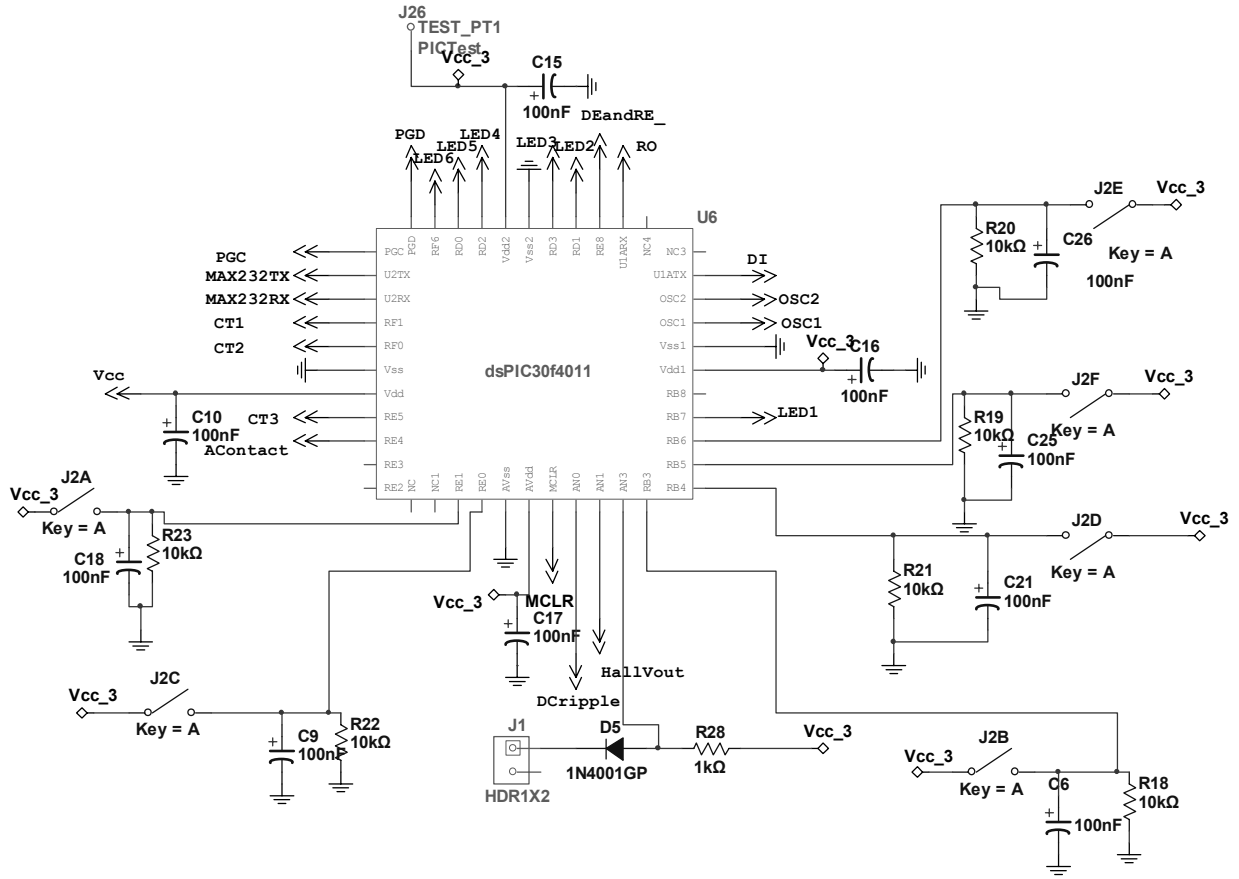


Fig. 6. Microprocessor

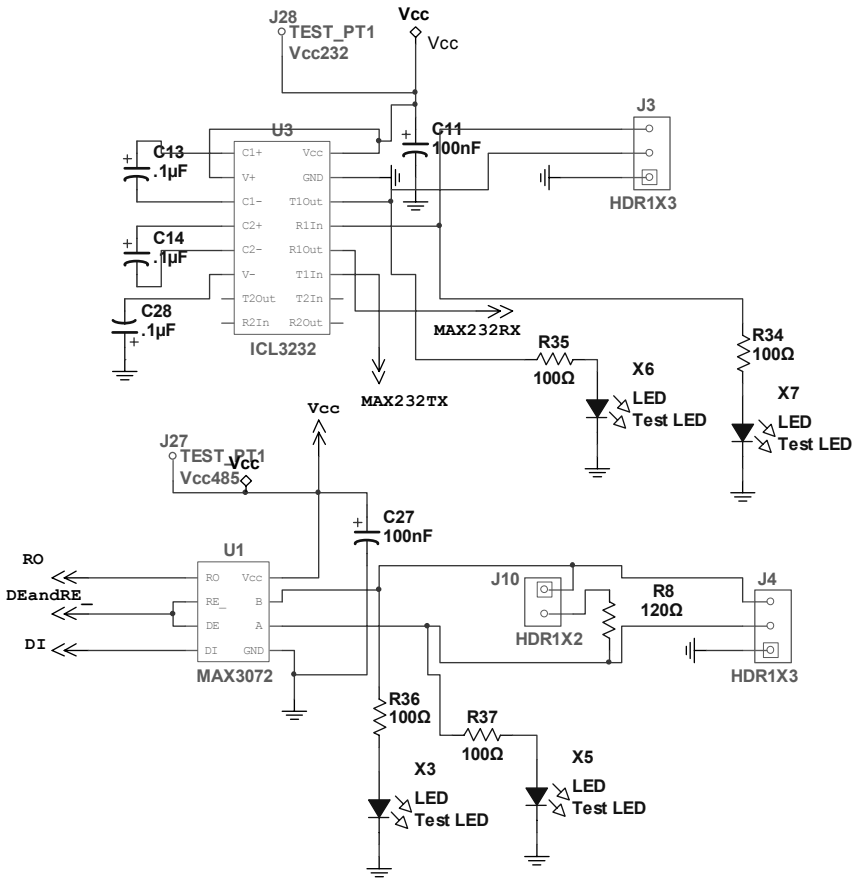


Fig. 7. RS-232 and RS-485 communication

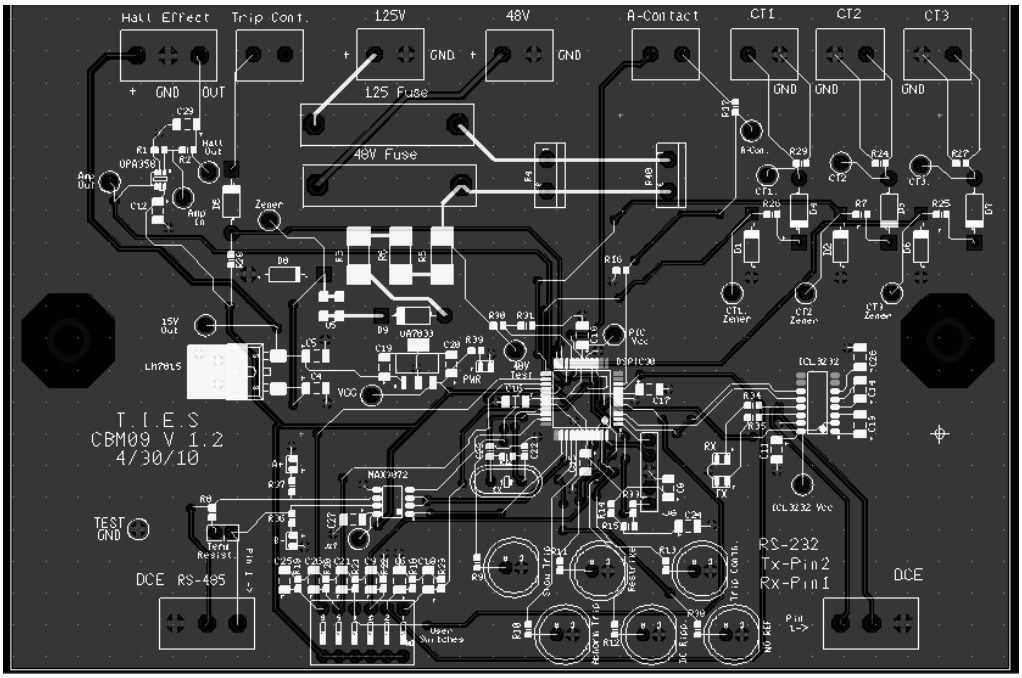


Fig. 8. PCB board layout

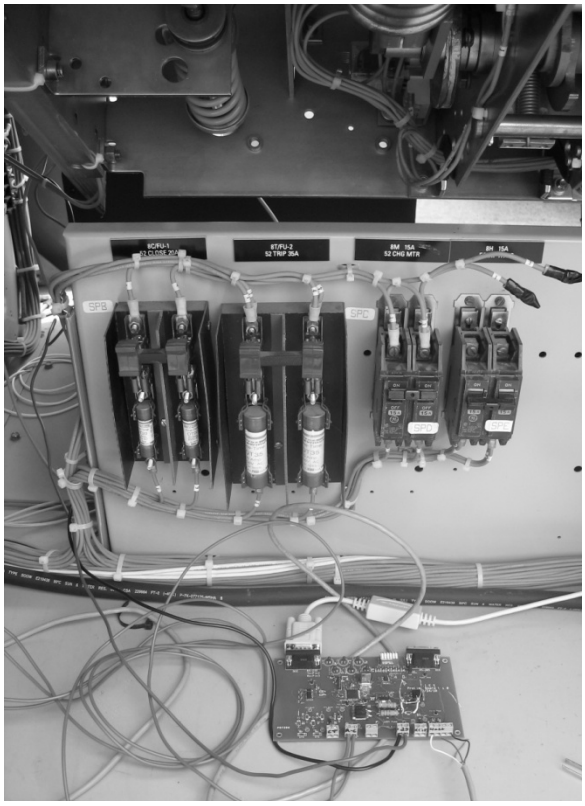


Fig. 9. Initial PCB tests

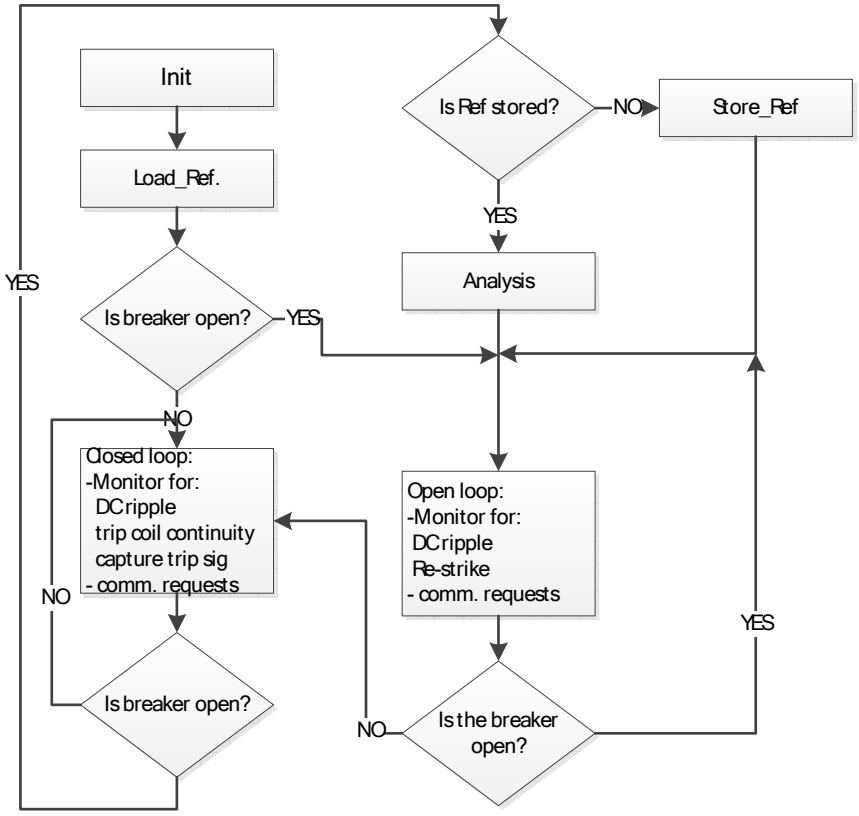


Fig. 10. Software top level flowchart

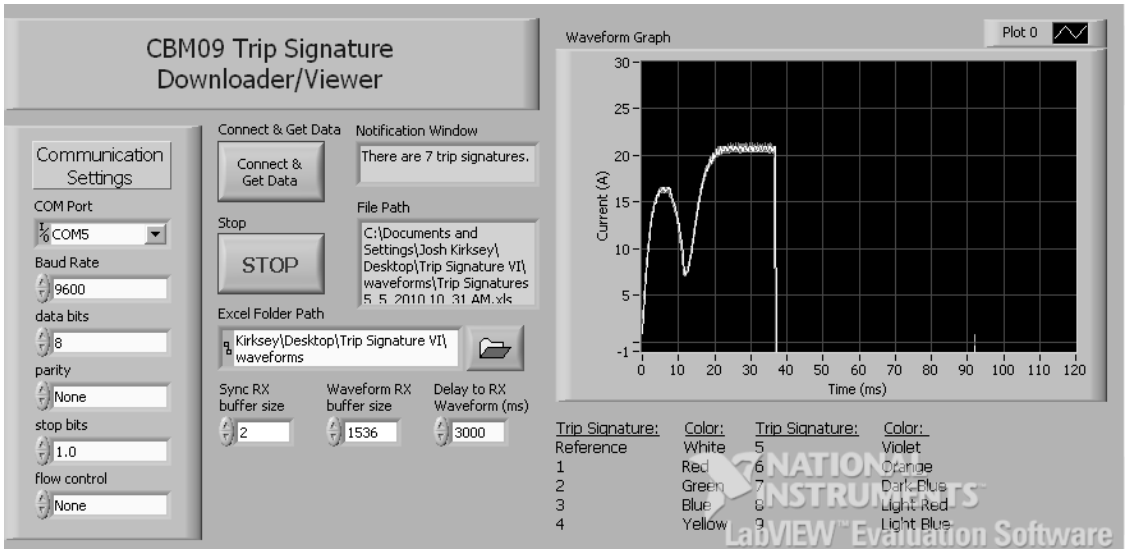


Fig. 11. Data acquisition

Fig. 12. Field test

Table 1. CBM09 Test Matrix

Tests \ Features	Monitor Trip Coil Status	Monitor Trip Signature	Monitor for Re-strike	Monitor for DC Ripple	Powered from 48V DC or 125V DC	Select either 3 or 5 Cycle Trip Time	Select between Sensitivities	Start/Stop Reference Capture	Clear Reference Waveform	Reset	Trip Signature Log	PC Software to Access and View Trip Signatures	RS-232 & RS-485 Communication	Size	Operational Temperature Range	Storage Temperature Range
CT Char. (3)			X													
Hall Effect Char.		X														
Power Module					X											
User Input (multi.)						X	X	X	X	X						
Phase Detector (3)			X													
a' Contact Detector	X		X													
DC Ripple Monitor				X												
Trip Coil Monitor	X	X														
PIC Input	X	X	X	X		X	X	X	X	X			X			
Trip Signature Proc.	X	X				X	X	X	X	X	X	X	X			
Re-Strike Proc.			X							X			X			
DC Ripple Proc.				X						X			X			
Data Storage		X						X	X	X	X	X				
LED Output	X	X	X	X	X	X	X	X	X	X			X			
RS-232 Comm.													X			
RS-485 Comm.													X			
Enclosure Testing														X		
Full System Testing	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 2. Power components cost comparison

Design 1			Design 2		
Part	**Qty.**	**Price**	**Part**	**Qty.**	**Price**
125 to 12 V DC/DC Converter	1	$77	Power Resistors	3	$12
			15 V Regulator	1	$1
48 to 12 V DC/DC Converter	1	$50	25 V 5W Zener Diode	1	$0.50
3.3 V Regulator	1	$1	3.3 V Regulator	1	$1
Total		$128	**Total**		$14.50

The Journal of Management and Engineering Integration
Vol. 3, No. 2, Winter, 2010

Applying Particle Swarm Optimization to the Multi-Level Lot Sizing Problem

Lotfi K. Gaafar
The American University in Cairo
gaafar@aucegypt.edu

Sarah M. El-Said
The American University in Cairo
smelsaid@aucegypt.edu

Abstract

This research investigates the potential of applying particle swarm optimization (PSO) to the multi-level lot sizing problem. The demand for the final product is assumed to be deterministic. The order size is constrained to multiples of a certain batch size. An all-ancestors cost-modified function is utilized to create interrelation between levels. The results obtained from the PSO are compared to those generated from a Wagner-Whitin (WW) algorithm. A 2^4 factorial experiment is designed to investigate the effects of four factors on the deviation of the PSO results from those of the WW. The factors are the demand pattern, product tree complexity, number of items in the product tree and length of the planning horizon. Comparisons were based on the difference between the costs of the order plans generated by the two algorithms and the frequency of each algorithm obtaining a lower cost. PSO outperformed WW in producing lower average cost plans in all experimental runs. All four factors and their interactions showed a significant effect on the difference between the costs of the order plans.

1. Introduction

One of the significant production problems is the lot sizing problem. The problem is basically concerned with finding the ordering quantities and their timings that minimize the total ordering and inventory holding costs for a given demand pattern of an item. The ordering cost denotes the fixed cost associated with the replenishment order regardless of its size. The inventory carrying cost includes the opportunity cost of the money (had it been otherwise invested) and the warehouse running cost. The cost of insufficient capacity in the short run (backordering cost) is the cost incurred as a penalty for not fulfilling the demand in the required period. The price of the item is usually neglected if it is constant over time [1].

The multi-level lot sizing (MLLS), also named multi-echelon problem or assembly problem, deals with an item that has several assembled components. Components are arranged in levels depending on precedence relationships to the item representing the end product. Unlike the single level lot sizing problem, planning for the MLLS problem is not only done at the final level for the end product but rather extends to the subcomponents as well.

Figure 1 represents the parts required to assemble a wireless mouse as an illustrative example for a MLLS problem. Figure 2 shows the product tree of the mouse assembly, where parts 7 and 8 are at level 2 two, subassembly 6 and parts 2, 3, 4 and 5 are at level one.

MLLS is directly related to material requirement planning (MRP) that translates the demand of the main item to the related demand of all its components. The internal relationship is represented by the bill of materials (BOM).The difficulty of the MLLS problem arises from the complexity of internal demand of several components with different schedules and fixed batch sizes. MRP is not a tool for making production decisions; instead production decisions are more of inputs to the MRP to function properly. MRP provides a feasible solution but not necessary the optimal to the multi-echelon lot sizing problem [2].

According to Dellaert *et al* [3], the MLLS problem can be divided into three categories based on the interrelationships between the items on the product tree as follows:

1- Pure assembly structure: each item has at most one immediate successor and any number of predecessors where the

predecessor item is a component assembled to the successor item. This is depicted in figure 3a, item 3 has one immediate successor (item 1) and two predecessors (items 4 and 5).

2- Pure abroscent structure: each item may have any number of immediate successors but at most one immediate predecessor as shown in figure 3b.
3- General structure: any item can have any number of successors and predecessors as shown in figure 3c.

1.1 Solution Approaches for the MLLS Problem

Approaches for solving the mathematical models such as linear programming, integer programming and relaxed versions of integer programming have been addressed in the literature [2, 4, 5, 6, 7]. Several heuristic methods were developed to solve the dynamic single level lot sizing problem. Although some heuristic methods offer near optimal solutions, these solutions are reached with less computations, complexity and effort. The most frequently used methods for solving this problem are the Silver-Meal heuristic and the Wagner Whitin algorithm.

In most cases solving the MLLS problem utilizing heuristic methods means repeating the application of the heuristic method for every item in the product tree. Dellaert and Jeunet [4] defined four heuristic strategies used to solve the MLLS problem; among which is the cost modification with single level single pass. The main aim of this method is to create an interrelation between the multi-levels while using single level pass. To incorporate the fact that an order in one level will impact the orders in the lower levels and thus the total costs, cost modification adds a fraction of the ordering cost of the subcomponents to the ordering cost of their successors.

Dellaert and Jeunet [4] addressed the multi-level lot sizing problem with no capacity constraints and a time invariant cost structure. They further extended the WW algorithm proposing a randomized cost-modified version of WW (CMWW) as well as a randomized cost-modified version of Silver-Meal (CMSM). They also developed randomized versions of other heuristics such as Grave's multi pass method, Bookbinder and Koch, Heinrich and Schneeweiss. Results showed that the cost modification procedure improves the performance of the previously mentioned heuristics. The randomized all-ancestors cost-modified Graves was reported as the best multi pass technique while the randomized all-ancestors CMWW is generally the second best technique and the first best single pass technique [4].

Recent literature emphasizes the power of meta-heuristics in solving complex combinatorial problems. Similar to heuristics, meta-heuristics offer a tradeoff between computational efforts and solution quality in terms of finding optimum solutions. Meta-heuristics have been applied widely in the field of dynamic lot sizing. They are often recommended for extensions of the lot sizing problem as they can be tailored to the problem; particularly when there is no heuristic available for solving the problem [8]. Meta-heuristic methods such as genetic algorithm (GA) and simulated annealing (SA) have provided promising results in solving the MLLS [3, 4, 9, 10, 11, 12]

PSO is another Meta heuristic method based on iterative population search. It was first introduced by Kennedy and Eberhart [13] to solve continuous non-linear functions. Particle swarm is inspired by observing bird flocking or fish schooling. Kennedy and Eberhart simulated the approach members of the bird flock use to exchange information to find food supplies. A member of the flock controls its own velocity based on two factors; its own best previous experience and the best experience of the other group members.

PSO initializes a first set of randomly generated potential solutions (particles) and then searches for the optimum one iteratively. It finds an optimal or near optimal solution by following the best particle and comparing its performance to that of other particles [14]. Unlike other evolutionary methods PSO does not get trapped in local minima easily because members of the population are maintained throughout the search process till finally reaching the best solution.

Although PSO was originally designed to solve continuous problems, in real life many discrete problems are encountered and therefore, Kennedy and Eberhart [15] developed a discrete version of PSO. The literature shows that PSO has been successfully applied to various problems including power and voltage control, artificial neural network training, pattern recognition and fuzzy control [16]. Other applications of PSO in continuous and discrete versions were provided by Kennedy *et al.* [17].

Applying PSO to the dynamic lot sizing problem has been very limited in literature. Since lot sizing is a discrete problem in nature, binary particle swarm was applied by Tasgetiren and Liang [18] to find the order quantities. Using PSO for solving a deterministic time-varying fixed quantity, single level lot sizing problem with backorders was examined by Gaafar and Aly [19]. Results reported by Tasgetiren and Liang [18] as well as Gaafar and Aly [19] showed that PSO performed better than other Meta heuristic methods.

2. Methodology

The current research aims at exploring the potential application of particle swarm optimization on the multi-level lot sizing problem. It attempts to find an improved solution for a multi-level product tree with a pre-determined demand over a finite planning horizon through evaluating the PSO performance by comparing it to a cost-modified WW. The current work applies a new version of discrete PSO using an all-ancestors cost-modified function to the MLLS problem with fixed lot size. A continuous PSO algorithm is initialized and then discretized to allow the three ordering options (ordering- no ordering and backordering). Therefore, the current work fills a gap through investigating the capabilities of PSO in solving the MLLS problem. The results obtained from the current research will be compared to the results of an all-ancestors cost-modified WW program developed to suit the problem specification. The cost-modified Wagner-Whitin (CMWW) has been chosen as a benchmark because it has the best reported results for the MLLS problem [4]. In the investigated MLLS problem, the orders can only be placed in integer multiples of a fixed batch sizes. Due to this restriction the quantity ordered could be more than the required demand incurring inventory cost or less than the demand incurring a backordering cost. Therefore, a discrete version of PSO is utilized. To accommodate the three options: no order, ordering a lower bound of the desired quantity or ordering an upper bound of the desired quantity as dictated by the fixed batch size.

The total cost of the plans developed by both algorithms will be compared through a designed experiment. Four factors will be tested: demand pattern, product tree complexity, number of items in the product tree and number of periods in the planning horizon. Other factors such as ordering cost, holding cost and backordering cost are randomized. The batch size is held constant throughout all runs.

2.1 Problem Assumptions

- The planning horizon is finite.
- The demand of the end product is defined along the planning horizon.
- The supply of each item in the product tree is infinite.
- No external demand for an intermediate item is allowed.
- Orders are delivered at the beginning of each period.
- Carrying costs are calculated based on the ending inventory at each period.
- No joint order discounts (i.e., the unit cost does not depend on the quantity ordered).
- No restriction on storage time (i.e., an item can be stored for several periods).
- The product has a maximum of three levels (including the zero level) with one end product at level zero.
- Setup costs and holding costs are defined along the planning horizon and do not change with time (i.e., costs are not affected by inflation or deflation of prices along the planning horizon).
- Orders are placed in integer multiples of a fixed batch size, where each item has its own batch size. For example, if the batch size for a given item is 7 then the number

of units ordered is either 7 or one of its integer multiples.

- Backordering is allowed to accommodate shortages due to ordering a lower bound of the batch size.
- An order must be placed in the first period of the planning horizon.
- The plan starts with zero inventory.
- Lead time is assumed to be zero.

The current structure of the PSO is different from both the continuous and the binary versions of PSO developed by Kennedy and Eberhart in two main aspects which are the particle representation and the particle update. Binary PSO was used to solve the lot sizing problem without batch size limitation. The particle is composed of a string of binary genes (bits) with a length that is equal to the planning horizon, where 0 denotes a period with no order and 1 denotes a period where orders take place. Binary PSO can only represent the two options of ordering and not ordering. On the other hand, if orders are placed in multiples of a fixed batch size, the binary PSO does not determine if an upper or a lower bound batch size is ordered. For example, if an item is ordered in batches that are equal to 6 units, and if the required demand for a group of periods is 17 units, an order quantity has to be either 12 or 18 units. To accommodate exploring those two options, a "0 1 2" is utilized in the current research. In "0 1 2" code, a zero denotes a no order period, 1 denotes a period with a lower multiple batch size order (12 in the example above) and 2 denotes a period with an upper multiple batch size order (18 in the same example).

2.2 Particle Representation

A string consisting of *N* genes is initiated. The number of genes *(N)* is equal to the number of periods in the planning horizon. The genes are assigned values that are $\in$ {0, 1, 2}. For Example, the highlighted PSO chromosome in the third row of Table 1 consists of 8 genes indicating that the planning horizon consists of 8 periods. The cumulative demand of periods 1, 2 and 3 is 51 units. If orders can only be placed in multiples of 6 units, the order quantity could either be 48 or 54 units. In period 1, the gene value is 1; therefore a lower bound of the order that is placed is equivalent to 48 units. No orders will be placed in periods 2, 3, 6 and 8 because they have a gene value of zero. The backorder of 3 units is added to the demand of period 4 for a total of 5 units. The gene value is 2 meaning that an order of 6 units is placed in period 4.

Given a certain demand pattern, a "0 1 2" PSO string is generated. The order quantity is calculated utilizing the order strategy suggested by the generated PSO string. Calculations of ordering quantities, beginning and ending inventories are carried out as presented in Table 1. Assuming a batch size of 6 units and since the quantity ordered in period 1 is less than the required demand; a backorder of 3 units is placed. Accordingly the net requirement in period 4 is modified to 5 units. To calculate the total cost, the ordering cost is assumed to be \$100, holding cost and backordering cost per unit are assumed to be \$0.5 and \$4 respectively.

2.3 Particle Update

The particle's position is updated by updating the positions of the genes within the particle. A gene's position is updated through calculating the gene's velocity from Eq.1. The velocity of any gene is a non-integer number that is limited to the range between [-3,3] inclusively. The position of the genes within the particle is then updated using Eq.2.

$$v_{k+1}^{i,j} = v_k^{i,j} + c_1 * (p_k^{i,j} - x_k^{i,j}) + c_2 * (p_k^{g,j} - x_k^{i,j}) \quad (1)$$

$$x_{k+1}^{i,j} = x_k^{i,j} + v_{k+1}^{i,j} \quad (2)$$

Where:

$v_k^{i,j}$: Velocity of bit *j* of particle *i* in iteration *k*

$x_k^{i,j}$: Position of bit *j* of particle *i* in iteration *k*

$v_{k+1}^{i,j}$: Velocity of bit *j* of particle *i* in iteration *k+1*

$x_{k+1}^{i,j}$: Position of bit *j* of particle *i* in iteration *k+1*

$p_k^{i,j}$: Position of bit *j* of best position reached by particle *i* in iteration *k*

$p_k^{g,j}$: Position of bit j of best global position g reached through iteration k

c_1 : Local drag factor, $c_1 \sim U[0,1]$

c_2 : Global drag factor, $c_2 \sim U[0,1]$

After updating the particle position, the values of the genes become non-integers as well. A flooring function is used to round down the absolute gene values to the closest integer $\in$ {0, 1, 2}, thus obtaining a "0 1 2" value again. For example, the position string given in figure 4 (a) is floored to the string in figure 4 (b).

The PSO algorithm is designed to optimize each level of the MLLS separately while keeping correlation between the levels through the cost-modified function using the pseudo code provided in figure 5.

2.4 The Cost-Modified Function

This research attempts to solve the MLLS problem through level-by-level optimization. Placing an order in one level triggers orders in lower levels. To create a relationship between levels, a cost-modified function will be utilized so that the best attainable solution is reached for every item and results are passed to the lower level [4].

Following the concept that level 0 represents a finished product, items are numbered sequentially starting from 1 (top to bottom and left to right) as in figure 2. Item 1 is the final item and other items are predecessors for it. The cost modified function adds to the cost of ordering of an item a percentage of the ordering cost of each of its predecessors. The cost modified function is not utilized in the last level because items in this level have no predecessors.

Dellaert and Jeunet [4] presented the mathematical formulation of the cost-modified function as follows:

$\beta_{i,j}$ is the successor based factor, representing the share that item i is responsible for in item j orders where item j is the predecessor for item i.

$$\beta_{(i,j)} = \frac{1}{\Gamma(j)} \qquad (3)$$

$$S_i = s_i + \sum_{j \in \Gamma^{-1}(i)} T(i,j,t)\beta_{i,j} s_j \qquad (4)$$

S_i : The total ordering cost for item i after cost-modification.

s_j : The ordering cost for item j

t: Time period where an order takes place $t \in \{1,2,3,\ldots,N\}$

l_i: Lead time for item i

$\Gamma^{-1}(j)$: Number of predecessors of item j

$\Gamma(j)$: Number of successors of item j

$T(i,j,t)$: A binary value that is set to 1 if ordering item i would place a new demand on item j in period t-l_i and 0 otherwise.

The CMWW algorithm was chosen as a benchmark for the current research since its randomized version was cited in literature as the most efficient single path technique to solve the MLLS (Deallart and Jeunet 2003). The current research uses the original single level algorithm with the addition of a cost-modified function. The original WW algorithm evaluates different options of planning. It sums up the orders of several periods until it is more economical to place a new order because the incurred inventory cost is becoming more expensive than placing a new order with a new ordering cost [1].

The WW algorithm guarantees optimality for single level lot sizing problems only. Similar to the work of Dellaert and Jeunet [4], the current research utilizes a WW algorithm for each item of the three levels of the problem. The outputs of one level are the input demand for the following level. After finding the best solution for each level, the total cost is calculated by summing up the costs of all items in the product tree.

The original algorithm does not incorporate fixed batch sizing. Similar to the work of Shittu [20], the current work incorporates batch size limitation. This is accomplished by executing a binary WW then switching the genes of the binary WW solution that have the value of "1" to the value of "2" randomly. This process is repeated 2000 times for each WW solution and the best chromosome is used as the final WW solution.

2.5 Data Generation

The current work deals with two sizes of the problem:

1- Small: 10 items and a planning horizon of 12 periods.
2- Large: 15 items and a planning horizon of 30 periods.

Each of the two problem variants is tested with two different demand patterns (constant and random). A constant demand means that for the tested problem instance the demand is constant over the planning horizon while a variable demand changes along the planning horizon. For the constant demand, a random number is generated form a discrete uniform distribution between 1 and 1000. The number is used as the demand for all periods. For the variable demand, a random number is generated from a discrete uniform distribution between 0 and 1000, for each period in the planning horizon. Two different product tree complexities (pure assembly and general structure) are used. The ordering, carrying and backordering costs are randomly generated integers following a uniform distribution. The range of the ordering cost is [100, 1000], and the carrying cost is [1, 10] while the backordering is [1, 50]. The batch size is fixed to 12 units for all items. The relationship between items in the product tree is represented by an immediate successor binary matrix that is generated randomly.

2.6 Experimental Design

The results of the PSO and the WW are compared in a designed experiment to test the significance of different factors. Four factors are tested; demand pattern (A), product tree complexity (B), number of items in the product tree (C) and length of planning horizon (D). Each of these factors is tested at two levels as shown in table 2. Thus, a 2^4 factorial experiment is conducted yielding 16 experimental runs. The demand is tested at constant and random patterns while the complexity is tested at low and high levels where the low complexity means that each item has at most one successor and any number of predecessors. The high complexity indicates that items can have any number of successors at any higher level. The number of items in the product tree is either 10 or 15 and the planning horizon is either 12 or 30 periods. The 16 experimental runs are executed, each with 100 instances, totalling 1600 instances. Each instance is solved using PSO then WW. The significance of the four factors and their interactions is tested on the percentage deviation of the results of the PSO from those of the WW (dPSO). The frequency of reaching a better solution is also investigated to check whether the PSO or the WW is more likely to reach a better solution.

3. Results

The results of the 16 experimental runs are summarized in table 3. The results in the dPSO column represent the percentage deviation in the average cost between the PSO and WW based on 100 instances for each experimental run. Each 50 instances are averaged representing a replicate. The average of both replicates is the dPSO. A negative sign in the dPSO column means that the average of the costs generated by the PSO is lower than those generated by the WW by the given value. Thus, a negative sign in that column indicates that PSO performs better than the WW. Table 3 shows that PSO yielded lower average cost values in all 16 experimental runs, with a minimum average improvement of 0.9% and a maximum average improvement of 24.07% over the WW. fPSO is the number of times PSO hit a better solution than the WW while fWW is the number of times WW hit a better solution than the PSO. The shaded cells in the fPSO and fWW columns represent the algorithm with the better performance. For example, in run number 1, out of the 100 instances PSO reached a better solution in 73 instances and WW reached a better solution in 12 instances. Both algorithms gave the same solution in 15 instances. PSO had a higher frequency of reaching a better solution in 15 experimental runs, while WW had a higher frequency in one run only. Although fWW is higher than fPSO in run number 11, PSO had a lower average cost by 3.86%.

Figure 6 compares the performance of the PSO to that of the WW over the 1600 instances. The given curve represents the percentiles of solution deviation of PSO from WW. The graph shows that PSO performs better

than WW in 65% of the instances. PSO and WW yielded the same results in 20% of the instances and WW outperforms PSO in 15% of the instances. The maximum improvement of the PSO over the WW is 166.2% while the maximum improvement of the WW over the PSO is 44.6%.

Statistical Analysis is applied to the results obtained from the 16 experimental runs. The average results of the 16 runs in table 3 were tested for normality. The 16 points seemed to fit in a straight line therefore the results of the 16 experimental runs show a normal distribution with a P-value of 0.016. The current research tests the results of 1600 problem instances; each is solved with both PSO and WW. Each pair of observations, is taken under homogeneous conditions since it results from testing the same problem with the PSO and then with WW. Accordingly, the results of the paired instances are dependant requiring the use of a paired t-test to compare the results of the PSO against those of the WW. The paired t-test checks if the mean of the generated results by the PSO is less than that of the WW according to the following hypothesis:

$H_o: \mu_D = 0$ $H_1: \mu_D < 0$ $\mu_D = \mu_{pso} - \mu_{ww}$

H_o is rejected and it can be concluded that there is enough statistical evidence to state that $\mu_D < 0$ and accordingly $\mu_{pso} < \mu_{ww}$. The P-value is 0.0005.

The significant factors and interactions affecting the previously discussed Experimental Design were examined through ANOVA. The response analyzed is the deviation of the PSO results from the WW (dPSO). Table 4 summarizes the results obtained from the ANOVA. Four main factors, six two-way interactions, four three-way interactions and one four-way interaction are assessed. The four main factors show a significant effect on dPSO. On changing any of the four factors from low to high the average dPSO decreases indicating that the PSO performs better than the WW. Table 5 shows the factors and interactions affecting the dPSO.

4. Conclusions

This research aimed at exploring the potentials of PSO in solving the MLLS problem. A discretized cost-modified version of PSO algorithm was developed and tested over 1600 instances utilizing design of experiment. The results were compared against those obtained from a cost-modified WW algorithm. PSO performed better than WW 65% of the time and yielded a maximum improvement of 166%. Results of a paired t-test show that PSO results have a statistically lower average total cost than the WW. This research can be further extended to solve more realistic problems with more complexities. MLLS Problems with capacity constraints on storage area or minimum order quantity can be investigated. Investigating the effects of other factors such as batch size and lead time can be studied. Solving problems with lead time can be explored through applying a different strategy to obtain the "0 1 2" code. For example instead of using a cost-modified function to relate all levels together; one long chromosome can be used where each item represent a segment in the chromosome. Segments can be further divided into genes that are equal to the number of periods in the planning horizon [11].

References

[1] Silver, E. and Peterson, R. (1998). *Decision System for Inventory Management and production planning*, Third Edition, John Wiley & Sons.

[2] Steinberg, E. and Napier, H. (1980). Optimal multi-level lot sizing for requirements planning systems. *Management Science*, 26 (12), pp1258-1271.

[3] Dellaert, N., Jeunet, J. and Jonrad, N. (2000). A genetic algorithm to solve the general multi-level lot sizing problem with time-varying costs. *Int. J. Production Economics,* 68 (3), pp 241-257.

[4] Dellaert, N. and Jeunet, J. (2003). Randomized multi-level lot-sizing heuristics for general product structures. *European Journal of Operation Research,* 148, pp 211-228.

[5] Voros, J. (1995). Setup cost stability region for the multi-level dynamic lot sizing problem. *European Journal of Operation Research*, 87, pp 132-141.

[6] Voros, J. (2002). On the relaxation of multi-level dynamic lot-sizing models. *Int J. Production Economics,* 77, pp 53-61.

[7] McKnew, M., Saydam, C. and Coleman, B. (1991) An efficient zero-one formulation of the multilevel lot-sizing problem. *Decision Sciences*, 22, pp 280-295.

[8] Jans, R. and Degraeve, Z. (2007). Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches. *European Journal of Operation Research*. 177, pp 1855-1875.

[9] Kuik, R. and Salomon, M. (1990). Multi-level lot sizing problem: evaluation of simulated annealing heuristic. *European Journal of Operation Research,* 45(1), pp 25-37.

[10] Tang, O. (2004). Simulated Annealing in lot sizing problems. *Int. J. Production Economics,* 88, pp 173-181.

[11] Prasad, P. and Chetty, O. (2001). Multilevel lot sizing with a genetic algorithm under fixed and rolling horizons. *Int J. Advanced Manufacturing Technology,*18, pp 520-527.

[12] Ozdamar, L., Barbarosoglu, G., (1999). Hybrid heuristic for the multi-stage capacitated lot sizing and loading problem, *Journal of the Operational research Society,* 50, pp 810-825.

[13] Kennedy, J. and Eberhart, C. (1995). Particle swarm optimization. *In: Proceedings of IEEE international conference on neural networks*, 4, pp 1942-1948.

[14] Shi, X., Liang, Y., Lee, H., Lu, C. and Wang, Q. (2007). Particle swarm optimization-based algorithms for TSP and generalized TSP. *Information Processing Letters,* 103(5), pp 169-176.

[15] Kennedy, J. and Eberhart, R. (1997). A discrete binary version of the particle swarm algorithm. *In: Proceedings of IEEE conference on systems, man and Cybernetics,* 5, pp 4104-4109.

[16] Lian, Z., Gu, X., and Jiao, B. (2006). A similar particle swarm optimization algorithim for permutation flowshop scheduling to minimize make span. *Applied Mathematics and computation,* 175(1), pp 773-785.

[17] Kennedy J., Eberhart R. and Shi Y. (2001). "Swarm Intelligence", Morgan Kaufmann, San Mateo, CA.

[18] Tasgetiren, M. and Laing, Y. (2003). A binary particle swarm optimization algorithm for lot sizing problem. *Journal of Economic and Social Research*, 5(2), pp 1-20.

[19] Gaafar, L. and Aly, A. (2008). Applying particle swarm optimization to dynamic lot sizing with batch ordering. *International Journal of Production Research,* (to appear).

[20] Shittu, E. (2003). Applying genetic Algorithm to the deterministic time varying fixed quantity lot sizing problem. The American University in Cairo.

Figure 1: parts required to assemble a wireless mouse.

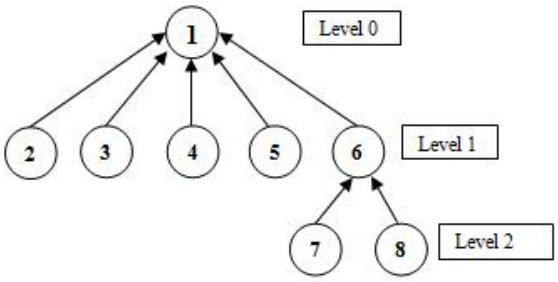


Figure 2: Product tree representing parts required to assemble the wireless mouse in figure 1

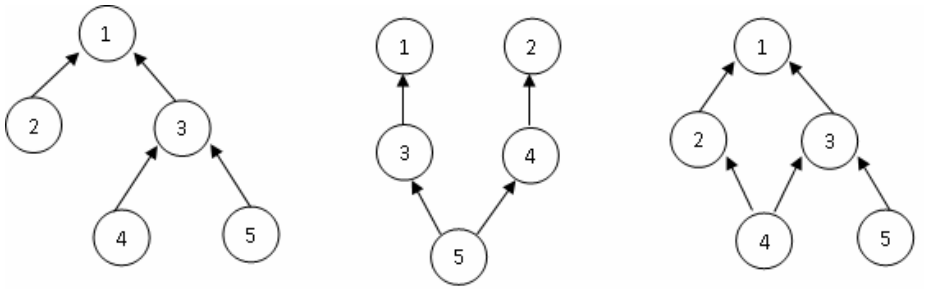


Figure 3: A schematic representation of the types of MLLS [3]

-0.5	*2.7*	*-1.3*	*0.8*	*4.1*	*0*

(a) Position string for "0 1 2" PSO

0	*2*	*1*	*0*	*2*	*0*

***(b)*Floored position string for a "0 1 2"PSO**

Figure 4: "0 1 2" PSO string flooring

Discretized PSO

k=0 (k is the iteration counter)
$kmax$=100 (number of iterations)
For i=1 to M (M=population size)
  For j=1 to N (N= number of planning periods)
  Initialize $x_k^{i,j}$
  $p_k^{i,j} = x_k^{i,j}$ (Initialize the "memory" for each particle)
  Initialize $v_k^{i,j}$
  End
  Evaluate objective function
End
p_k^g =Best particle found in $x_k^{i.}$
k=k+1
While $k < kmax$
  For i=1 to M
    For *j=1 to N*
    (calculate the speed of each gene)
    $v_{k+1}^{i,j} = v_k^{i,j} + c_1 * (p_k^{i,j} - x_k^{i,j}) + c_2 * (p_k^{g,j} - x_k^{i,j})$
    (calculate the new position of each gene)
    $x_{k+1}^{i,j} = x_k^{i,j} + v_{k+1}^{i,j}$
    Floor $x_{k+1}^{i,j}$
    End
  Evaluate objective function
  Find personal best $p_k^{i.}$
  Find the global best p_k^g
  End
k=k+1
End

Figure 5: The pseudo code used for the discretized PSO

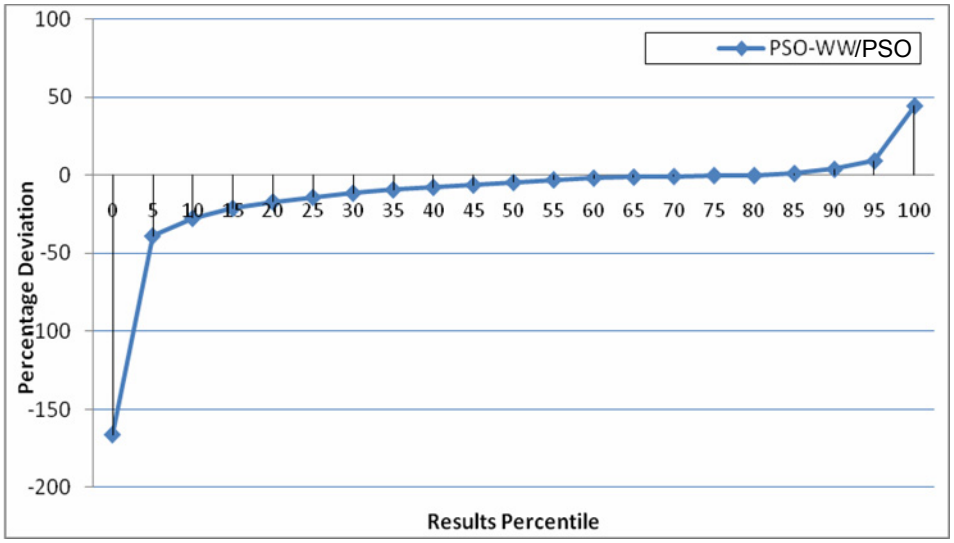


Figure 6: Percentile of deviation of PSO from WW

Table 1: Order quantity and cost calculations for a "0 1 2" PSO chromosome

Period	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Cost*
Demand	16	21	14	2	8	7	11	21	
PSO chromosome	1	0	0	2	1	0	2	0	
Net requirement	51	0	0	5	15	0	34	0	
OQ	48	0	0	6	12	0	36	0	4*$100=$400
BI	0	32	11	0	1	5	0	23	
EI	32	11	0	1	5	0	23	2	74*$0.5=$37
BO	0	0	3	0	0	2	0	0	5*$4=$20
Total Cost									**$457**

NR: net requirement BI: beginning inventory EI: ending inventory
BO: backordering OQ: order quantity
(*): net requirements after adding backorder

Table 2: Factors tested in the designed experiment

Factor	***Levels***	
	LOW	**HIGH**
Demand Pattern (A)	Constant	Random
Product tree complexity (B)	Assembly structure	General structure
Number of items in the product tree (C)	10	15
Length of planning horizon (D)	12	30

Table 3: Results of the factorial experiment

Run	FACTORS				RESPONSES				
	A	B	C	D	Replicate 1*	Replicate 2*	dPSO	fPSO	fWW
1	const	low	10	12	-3.38	-2.56	-2.97	73	12
2	random	low	10	12	-6.39	-4.73	-5.56	58	42
3	const	high	10	12	-2.96	-1.14	-2.05	69	19
4	random	high	10	12	-2.32	-4.27	-3.30	62	38
5	const	low	15	12	-10.05	-8.07	-9.06	80	5
6	random	low	15	12	-11.72	-8.87	-10.29	87	3
7	const	high	15	12	-8.53	-9.55	-9.04	83	5
8	random	high	15	12	-10.05	-9.31	-9.68	90	2
9	const	low	10	30	-3.02	-4.37	-3.70	55	45
10	random	low	10	30	-8.18	-7.88	-8.03	74	16
11	const	high	10	30	-3.49	-4.22	-3.86	49	51
12	random	high	10	30	-4.90	-6.21	-5.55	55	45
13	const	low	15	30	-0.90	-0.97	-0.90	100	0
14	random	low	15	30	-19.55	-21.22	-20.38	83	8
15	const	high	15	30	-20.44	-27.72	-24.07	81	7
16	random	high	15	30	-18.10	-19.01	-18.56	85	8

*based on an average of 50 instances

(a) Factors key

A	Demand Pattern
B	Complexity of tree
C	Number of Items
D	Length of planning periods

(b) Responses Key

dPSO	Deviation of PSO from WW
fPSO	Number of times PSO hit a better solution than the WW
fWW	Number of times WW hit a better solution than the PSO

Table 4: ANOVA results

Factors	*Degrees of freedom*	*Sum Square*	*Mean Square*	f_o	*P*
A	1	82.42	82.42	31.08	0.000
B	1	28.82	28.82	10.86	0.004
C	1	561.32	561.32	211.68	0.000
D	1	137.28	137.28	51.77	0.000
AB	1	109.05	109.05	41.12	0.000
AC	1	4.4	4.4	1.659	0.215[a]
AD	1	25.36	25.36	9.564	0.006
BC	1	85.72	85.72	32.32	3.3E-5
BD	1	65.05	65.05	24.52	0.0001
CD	1	43.29	43.29	16.32	0.0009
ABC	1	58.26	58.26	21.97	0.0002
ABD	1	82.39	82.39	31.07	4.2E-5
ACD	1	12.17	12.17	4.58	0.0479[a]
BCD	1	55.59	55.59	20.96	0.0003
ABCD	1	66.55	66.55	25.09	0.0001
Error	16	42.43	2.65		
Total	31	1460.12			

a Insignificant effects (at $\alpha = 0.01$), but are likely masked out by higher order interaction

Table 5: factors and interactions affecting the dPSO

	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
dPSO	√	√	√	√	√	√*	√	√	√	√	√	√	√*	√	√

√ indicates that the given factor or interaction significantly affects the dPSO (α=0.01).
* indicates that the given interaction has an insignificant effect that is probably masked out by higher order interaction.