

School of Engineering and Computing

**Computing and Information
Systems Journal**

Vol 21, No 3, 2017

Edited by Abel Usoro



© University of the West of Scotland, 2017

All authors of articles published in this journal are entitled to copy or republish their own work in other journals or conferences. Permission is hereby granted to others for the publication of attributed extracts, quotations and citations of material from this journal. No other mode of publication or copying of any part of this publication is permitted without the explicit permission of the University.

Computing and Information Systems is published normally three times per year, in February, May, and October, by the University of the West of Scotland.

From the next issue the editorial address is Dr Abel Usoro, School of Computing, University of Paisley PA1 2BE; tel (+44) 141 848 3959; fax (+44) 141 848 3542, e-mail: cis@uws.ac.uk or abel.usoro@uws.ac.uk.

Editorial Policy

Computing and Information Systems offers an opportunity for the development of novel approaches, and the reinterpretation and further development of traditional methodologies taking into account the rate of change in computing technology, and its usage and impact in organisations.

Computing and Information Systems welcomes articles and short communications in a range of disciplines:

- Organisational Information Systems
- Computational Intelligence
- E-Business
- Knowledge and Information Management
- Interactive and Strategic Systems
- Engineering
- E-Learning
- Cloud Computing
- Computing Science

The website for Computing and Information Systems is <http://cis.uws.ac.uk>

Implication of Variation in Ant Colony Optimization Algorithm (ACO) Parameters on Feature Subset Size

Ronke S. Babatunde, Rafiu M. Isiaka, Oyeranmi J. Adigun and
Akinbowale N. Babatunde

^{1,2,4}Department of Computer Science, Kwara State University, Malete, Nigeria.

³Department of Computer Technology, School of Technology, Yaba College of Technology, Lagos, Nigeria.

ronke.babatunde@kwasu.edu.ng, abdulrafiu.isiaka@kwasu.edu.ng, ranmiadigun@gmail.com,
akinbowale.nathaniel@kwasu.edu.ng

ABSTRACT

Purpose: The purpose of this research is to apply ant colony optimization (ACO), a nature inspired computational (NIC) technique to achieve optimal feature subset selection, for the purpose of data dimensionality reduction, which will remove redundancy, provide a reduced storage space, improve memory utilization and subsequently, classification task.

Design/Methodology/Approach: The ACO feature subset selection process was filter-based, accomplished by using correlation coefficient as the heuristic information η which guides the search to determine the pixel (feature) attractiveness. The intensity of the pixel represents the pheromone level τ . Also, ρ , ϕ , α and β are key parameters which were tuned to various arbitrary values at each iteration to determine the best combination of parameters which can result in a reduced feature subset (window size). The procedure was simulated in MATLAB 2012.

Findings: The study has been able to demonstrate that experimental results obtained by tuning the parameters of ACO at each run gave a reduced feature subset (window size), at parameter settings of α (0.10), β (0.45), ρ (0.50), ϕ (1.0) which was salient for efficient face recognition thereby demonstrating the effectiveness and superiority of filter-based over wrapper-based feature selection.

Research Limitation: The major limiting factor in this research was time and materials which hindered the further testing and implementation of innovative computational intelligence.

Originality/Value: This research work is valuable because it gives the implication of the variation of the ACO parameters which determines the extent to which the optimal

solution construction that results in a reduced feature subset can be realized.

Keywords: *feature selection, Ant Colony Optimization, pattern recognition.*

Paper Type: *Implementation*

1. Background to study

The recent outburst of data set size, in terms of number of records as well as of attributes, has triggered the development of a number of big data platforms as well as parallel data analytics algorithms. At the same time though, it has pushed for the usage of data dimensionality reduction procedures (Mauricio and Helio, 2013). In a bid to simplify the amount of information used to describe a large set of data, features are selected from the data, serving as distinctive representative of its contents. In image analysis, for instance, features include curves, color, texture, edges, object shape, interest points, among others. These features usually are organized into an n-dimensional feature vector (Mauricio and Helio, 2013). Feature selection problem consists in making good predictions with as few variables/features as possible. The optimality of a feature subset is measured by an evaluation criterion (Kashef and Nezamabadi-pour, 2014). The main purpose of feature selection is to reduce the number of features used in classification while maintaining acceptable classification accuracy (Raymer *et al.*, 2000).

Feature selection is an important step in many pattern classification problems. It is applied to select a subset of features, from a much larger set, such that the selected subset is salient enough to perform the classification task (Rosaria, Iris, Aaron and Michael, 2014). Locating an optimal feature subset is generally tasking, making most

problems related to feature selection to be classified as non-deterministic polynomial (NP) time hard (Babatunde, Olabiyisi, Omidiora and Ganiyu, 2014). Feature selection reduces computational requirements, storage space, and cost of future measurements as well as improves prediction performance and data or model understanding (Guyon and Elisseeff, 2003).

The basic feature selection approaches include wrapper-based, filter-based and hybrid approaches. The latter combines both filter and wrapper. In a wrapper feature selection, a classification algorithm is used to estimate the performance of the selected subsets, wherein the classifier serves as an evaluation function. On the other hand, filter uses information derived from the features in the given data to determine the local importance of a given feature, measured using some evaluation criterion such as the F-score, Mutual Information (MI), Information Gain (IG), entropy measure, correlation coefficient and several other information evaluation functions. With this, the intrinsic properties of the data is been measured. The algorithm calculates a relevance score for each feature. Features having low relevance score are removed. (Hatua, 2014)

Hybrid methods refer to a wide variety of techniques in which training data are given to a learning machine, which returns a predictor and a subset of features on which it performs predictions. Feature selection is performed in the process of learning, which saves the trouble of a two-step induction process (Guyon and Elisseeff, 2003).

Most promising approaches to feature selection still suffer from a number of major challenges limiting their implementation and use in practice. Thus, optimal feature subset selection remains an open problem domain.

With the objective of solving optimization problems, several bio inspired algorithms have been proposed in the literature, which include genetic algorithms (GA) and swarm algorithms such as particle swarm optimization (PSO), ant colony optimization (ACO), Clonal Selection Algorithm (CSA), Bat Algorithm and Artificial Bee Colony. ACO, of all the numerous traditional computational intelligence paradigms, has a unique feature of constructive search via indirect communication with the environment (stigmergy).

The crux of this research is the use of ACO for feature selection using filter-based approach.

2. ACO metaheuristic

A metaheuristic refers to a set of algorithmic concepts defining heuristic methods which can be useful in solving extensive set of diverse problems. In other words, a metaheuristic is a general-purpose algorithmic framework that can be applied to different optimization problems with relatively few modifications. Optimization methods can be used in the process of feature selection to determine the most relevant subset of features from the data set while maintaining adequate accuracy rate represented by the original set of features (Mauricio and Helio, 2013).

2.1 Feature Subset Selection in ACO

Ant Colony Optimization (ACO) techniques are based on the behaviour of real ant colonies used to solve discrete optimization problems (Mehdi, Nasser, Mohammad, 2009). Naturally, real ants are capable of finding the shortest route between their nest and food source without the use of visual information and hence possess no global world model, adapting to changes in the environment. The deposition of pheromone is the main aspect which enables real ants to find the shortest routes over a period of time. Each ant probabilistically prefers to follow a direction rich in this chemical. Based on this idea, artificial ants can be deployed to solve complex optimization problems via the use of artificial pheromone deposition (Dorigo, Birattari and Stutzle, 2006).

This method is particularly attractive for feature selection as there seems to be no heuristic that can guide a search to the optimal minimal subset (of features) every time. Also, the ants can discover the best feature combinations as they proceed throughout the search space.

In ant colony optimization, the problem is dealt with by simulating a number of artificial ants moving on a graph that encodes the problem itself: as in the case of a digitized face image, each vertex represents a pixel (feature) and each edge represents a connection between two pixels. A variable called pheromone is associated with each edge and can be read and modified by ants. Ant colony optimization is an iterative algorithm, and

at each iteration, a number of artificial ants are considered. Each of them builds a solution by walking from pixel to pixel on the face with the constraint of not visiting any pixel that has already been visited in the walk (Mauricio and Helio, 2013). At each step of the solution construction, an ant selects the subsequent pixel to be visited according to a stochastic mechanism that is biased by the pheromone: when in pixel i , the same pixel is selected stochastically among the previously unvisited ones. At the end of an iteration, on the basis of the quality of the solutions constructed by the ants, the pheromone values are modified in order to bias ants in future iterations to construct solutions similar to the best ones previously constructed (Mehdi, Nasser, Mohammad, 2009). A pheromone value is associated with each possible solution component; that is, with each possible assignment of a value to a variable. Formally, the pheromone value $\tau_{i,j}$ is associated with the solution component $C_{i,j}$, which consists of the assignment $X_i = V_{j,i}$ where X_i is the pixel at node i and $V_{j,i}$ represents movement from pixel j to i .

The set of all possible solution components is denoted by C . Here, solution components are pairs of pixels to be visited one after the other. The solution construction is based on the probabilistic transition rule shown in equation 1

$$p_{i,j}^k(t) = \frac{[\tau_{i,j}(t)]^\alpha [\eta_{i,j}]^\beta}{\sum_{j \in J^k} [\tau_{i,j}(t)]^\alpha [\eta_{i,j}]^\beta}, \text{ if } i \in J^k \quad (1)$$

where J^k is ant k 's unvisited features, η_i is the heuristic desirability of choosing feature i , τ_i is the pheromone value at feature i , α determine the importance of pheromone value and β determines the importance of heuristic information. J_i^k is the neighborhood of ant k when in node i . α determines the extent to which pheromone information is used as the ants build their solution. β determines the extent to which heuristic information is used by the ants.

Figure 1 illustrates the setup of ACO algorithm based feature selection. The ant is currently at node a and has a choice of which feature to add next to its path (dotted lines). It chooses feature b next based on the transition rule, then c and then d . Upon arrival at d , the current

subset ($a; b; c; d$) is determined to satisfy the traversal stopping criteria.

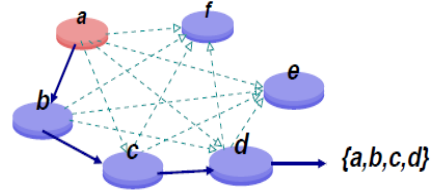


Figure 1: ACO problem representation for feature selection (Source: Jensen and Shen, 2005, pp 5-12)

A suitable heuristic desirability of traversing between features could be any subset evaluation function - for example, F-score criterion, mutual information measurement, information gain, entropy-based measure, fuzzy-rough set dependency measure, and correlation (Kashef and Nezamabadi-pour, 2014).

The heuristic desirability and pheromone factors are combined to form the so-called probabilistic transition rule, denoting the probability of an ant k at feature i choosing to move to feature j at time t :

The general ACO pseudocode proposed by Chen, 2011 is shown in Figure 2

Pseudocode

Set parameters, initialize pheromone trails
while termination condition not met do
 Construct Ant Solutions
 Apply Local Search (optional)
 Update Pheromones
End while

Figure 2: ACO Metaheuristics

The pheromone trail decay coefficient (ρ) and pheromone amount (φ) have an important impact on the performance of ACO (Mehdi, Nasser, Mohammad, 2009). The deposited pheromone, calculated using equation 2

$$\tau_{i,j}^k \leftarrow (1 - \rho) \cdot \tau_{i,j} + \rho \cdot \Delta(i,j)^k \quad (2)$$

where $\rho \in (0,1)$ is the pheromone update parameter and $\Delta(i,j) = \sum_{k=1}^n \Delta \tau_{i,j}^k$ and pheromone decay, calculated using equation 3.

$$\tau^{(n)} = (1 - \varphi) \cdot \tau^k + \varphi \cdot \tau^{(0)} \quad (3)$$

where φ is the decay coefficient, expresses the quality of the corresponding solution. φ simulates the pheromone evaporation. Evaporation becomes more important for more complex and non-linear problems such as facial feature selection. If $\rho = 0$, this implies no evaporation and the algorithm does not converge. If pheromone evaporates too much (i.e. a large ρ is used), the algorithm often converged to suboptimal solutions for complex problem. In many empirical problems, it is difficult to select the best combination of the parameters without a trial-and-error process. α and β are also key parameters in ACO for feature selection.

2.2 Related works

The parameter settings affect the impact of adaptation methods on the performance of the ACO algorithm. These determine the behaviour of ants and are critical for fast convergence to near optimal solutions of the given problem instance (Stutzle, 2002). The use of adaptation methods is motivated by the fact that an instance-optimal parameter setting always obtains better results than any other setting. ACO uses random input uniformly distributed in the interval $[0,1]$ (Dorigo, Birattari and Stutzle, 2006).

The most critical parameters in ACO algorithm are the influence of the pheromone trails α , the heuristic information β , pheromone update factor ρ and pheromone decay coefficient φ (Stutzle et al, 2010). The choice of value of ρ is directly related to the global search ability and the rate of convergence which are the two performance indices. α represents the strength of randomness in the search process. The bigger the value of α , the weaker the random performance, making the algorithm to be trapped in a local optima. β represents the strength of apriority and certainty factors in the process of searching. The bigger the value, the higher the possibility that the ants choose the local best result. High value of φ implies high pheromone decay rate which makes the algorithm fast in solution construction (Stutzle et al, 2010).

Hamidreza, Karim and Sayyed (2007) developed a feature selection using ant colony optimization and carried out a comparative study of the proposed algorithm with Genetic Algorithm on Face Recognition System. Using the wrapper

approach, the method uses classifier performance and the length of selected feature vector as heuristic information for ACO to select the optimal feature subset without the priori information of features. This approach was simulated on ORL and it yielded higher accuracy and a bit low computational complexity on ACO than GA.

In Zhou *et al.* (2008) a GA based wrapper method for classification of hyper spectral data using SVM classifier was presented. The GA was used to optimize both the feature subset of hyper spectral data in a wrapper approach as well as SVM kernel parameters simultaneously. Comparison of the optimized results and the un-optimised results showed that the GA-SVM method could significantly improve the classification accuracy.

Imani *et al.* (2012) integrated ACO and GA in a novel hybrid feature selection and support vector machine as classifier. The work used the combined algorithm to select the optimal feature subsets from features of Persian font images. Although the method gave a much reduced feature subset compared to the original, the time complexity to arrive at the result was high.

Pali and Bhaiya (2013) proposed a method of GA based neural network for feature selection. In the method which consists of four stages, PCA was used for dimensionality reduction, LDA was used for feature extraction, and GA was used for feature selection which wraps around Back Propagation Neural Network (BPNN) for evaluating feature subset as well as classification of face images. The work yielded high accuracy but with enormous execution time. There also exists several work combining filter and wrapper feature selection approaches sequentially.

Vidyavathi and Ravikumar (2008) presented a novel feature selection method based on two stage analysis of fisher ratio and mutual information. These two stage analyses were carried out in the feature domain to reject the noisy feature indexes and select the most informative combination from the remaining. The effectiveness of this method was evaluated by the SVM classifier and results obtained showed that the method can get higher prediction accuracy with small number of features. Argawal *et al.* (2010) presented a face recognition based on information theory approach of coding and

decoding the face image. PCA was employed for feature extraction and feed forward back propagation Neural Network for recognition of faces from ORL face database.

3. Research Approach

This section describes the method employed as well as the experimental setup of this research, the computational resources used to implement and evaluate the proposed feature selection method. The data set used was a locally acquired face database (LAFD) collected at the Kwara State University, Malete, Nigeria. All the experiments were conducted on a computer with Intel Core i7-2600 3.4 GHz and 4-GB RAM.

Prior to feature subset selection using ACO, the acquired face images were preprocessed for uniformity and normalization. At first, geometric normalization was done to help the comparability of the face image. Using Adobe Photoshop image editing software, cropping was done so as to retain only the face region, with the extinction of areas such as ear and fore-head without distortion. Then, photometric normalization was carried out on the LAFD images to eliminate illumination effects, improve contrast, enhance visual quality and to obtain a uniform histogram of the images. The local contrasts of the enhanced images are observed to have improved based on visual quality and the features on the face are more apparent.

Additionally, Feature extraction, which involves transformation of data to represent samples accurately in a lower dimensional feature space, was done using Local Binary Pattern (LBP) texture analysis. This was done to extract 2 dimensional distributions of local binary pattern and local contrast measurement. This enables texture information to be extracted from the face. The output here gives a suitable representation for ACO algorithm to work on the pattern for selecting optimal feature subsets. Therefore, the feature selection procedure of ACO is the construction of optimal solution from the pattern obtained in the feature extraction process.

The output of LBP was converted to image data matrix to become input into ACO. The image data matrix was created by converting the texture descriptors to double data format in order to set pixels (features) in a double array so that the matrix of each image can be obtained easily.

Pheromone matrix and other ACO parameters were initialized and the construction of solution begins by selecting pixels from the image data matrix based on feature-feature correlation and updating the pheromone matrix. The ants move randomly over the face in a clique to update values in the pheromone matrix (Babatunde, Olabiyisi, Omidiora and Ganiyu, 2015)

The size of the pheromone matrix for this experiment varies with a different resolution of the cropped image, arbitrarily chosen as 40*40, 50*50, 60*60, 70*70, 80*80, 90*90 and 100*100. Global pheromone update was done on the pheromone matrix within the construction step of each ant in one iteration using Equation (2). If all K ants in an iteration have moved for solution construction, then local pheromone update was done, else solution construction continues until all ants in that iteration have moved. Local pheromone update was done using Equation (3) on the pheromone matrix after an iteration was completed so that future ants in the next iteration could explore other paths and not construct the same solution. When the set number of iteration is reached the optimal feature subsets is obtained from the pheromone matrix and the feature selection procedure ends, else solution construction continues in a next iteration.

The heuristic desirability, η used was dynamic, i.e. obtained by determining the correlation between pairs of pixels at each ant move using Pearson product moment correlation (PPMC) method. Once two pixels are correlated, one of it is chosen, and considered sufficient enough to represent the facial information at that region. Once this is done to every pixel in the image, a subset of pixels is obtained which represents the optimal set of features on the face image. These were obtained by storing the pixels selected in a variable which gives the number of rows and columns at the end of the operation. The size of feature input is the dimension of the faces used in this experiment, which were 1600, 2500, 3600, 4900, 6400, 8100, and 10000 pixels for 40*40, 50*50, 60*60, 70*70, 80*80, 90*90 and 100*100 resolutions respectively.

The behaviour of ACO algorithm depends strongly on the values given to its parameter. Varying the parameters at computation time can enhance the performance of the algorithm. The quality of solution obtained by ACO was analyzed

over subset size for the specific parameter values. The goal is to identify where the various combinations of parameters give reduced feature subset compared with the original feature input.

4. Discussion of Result

In the experimental setup the values of α , β , ρ and ϕ were varied and based on the values of the parameter settings, the exploration and exploitation for effective solution construction was achieved by comparing the size of feature subset output with the original input features. The parameter value uses random values in interval $[0,1]$. For the experiment, arbitrary intervals were chosen and α was varied at intervals of 0.025, β at 0.15, ρ at 0.125 and ϕ at 0.25. Different values of

increment of the parameters at the intervals were varied and feature subset size obtained for several ACO runs. The result and average feature size selected are as shown in Table 1. It was observed that the feature subset size obtained compared to the original feature input gave the least window sizes of pixel resolutions at parameter settings of α (0.10), β (0.45), ρ (0.50) and ϕ (1.0), which indicates that the optimal subset of features was obtained at the combination of these parameter settings. It was also observed from the result as shown in the table that the values of the parameter below these settings and above these settings gave a higher window size compared to the original feature input.

Table 1: Window Size (Feature Subset) Output for various parameter tuning

Image Resolution (pixels)				40*40	50*50	60*60	70*70	80*80	90*90	100*100
Feature Input				1600	2500	3600	4900	6400	8100	10000
Parameters										
α	β	ρ	Φ	Average Feature Subset Selected (Window Size)						
0.025	0.15	0.125	0.25	26x28	29x28	32x34	38x34	49x51	56x59	66x69
0.050	0.25	0.250	0.50	22x24	30x31	27x22	37x39	50x54	61x67	68x70
0.075	0.35	0.375	0.75	20x23	28x26	26x27	33x36	41x45	53x59	68x76
0.100	0.45	0.500	1.0	11x13	17x16	19x16	24x23	27x29	32x26	39x43
0.125	0.55	0.625	1.25	23x25	32x30	29x31	43x41	51x53	61x64	72x77
0.150	0.65	0.750	1.5	28x26	33x31	30x34	48x50	54x57	63x62	69x73
0.175	0.75	0.875	1.75	31x 33	35x39	39x43	42x48	53x58	60x67	72x79

5. Conclusion and Future work

This research has carried out a filter-based feature subset selection using Ant colony optimization technique which employed correlation between pairs of pixels to determine the desirability of a pixel on a human face. The result obtained suggests that the method was suitable to reduce the spatial dependency among pixels on a face, which could lead to a reduced feature subset size. A faster recognition result can be obtained using a reduced feature subset (window size). Future work shall focus on exploration of other

innovative computational intelligence such as superbug algorithm (SuA), Cuckoo Search, and Galaxy based search algorithm.

References

Dorigo, Marco, Blum, Christian. (2005). Ant colony optimization theory. A survey. *Theoretical Computer Science*. Vol. 344, pp 243-278. Available: www.elsevier.com/locate/tcs. Accessed on 24th September, 2016.

- Omidiora, E. O, Fakolujo O. A, Ayeni R. O, Olabiyisi S. O, Arulogun O. T. (2008). Quantitative evaluation of principal component analysis and fisher discriminant analysis techniques in face images. *Journal of Computer and its Applications*. Vol. 15, Issue 1, pp 22-37.
- Babatunde, R.S., Olabiyisi S.O., Omidiora E.O. and Ganiyu R. A. (2014) Feature Dimensionality Reduction using a Dual Level Metaheuristic Algorithm. *International Journal of Applied Information Systems (IJ AIS)*, Vol. 7, Issue 1, pp 49-52.
- Babatunde, R. S., Olabiyisi S.O., Omidiora E.O. and Ganiyu R. A. (2015). Local Binary Pattern and Ant Colony Optimization Based Feature Dimensionality Reduction Technique for Face Recognition Systems. *British Journal of Mathematics & Computer Science*, Vol. 11, Issue 5, pp 1-11.
- Hamidreza, R. K, Karim F and Sayyed M. T. (2007). Feature Selection Using Ant Colony Optimization (ACO). A New Method and Comparative Study in the Application of Face Recognition System. In Perner, P. (Ed.) *ICDM, LNAI 4597*, pp 63–76, © Springer-Verlag Berlin Heidelberg
- Rosaria Silipo, Iris Adae, Aaron Hart and Michael Berthold (2014). Seven Techniques for Dimensionality Reduction. White Paper by KNIME.com AG. pp 1-21.
- Ahmed, Al-Ani, M. Deriche and J. Chebil. (2003). A new mutual information based measure for feature selection. *Intelligent Data Analysis*, Vol. 7, pp 43-57.
- Zhuo, Li., Zheng, Jing, Wang, Fang, Li, Xia., Ai, Bin., Qian, Junping. (2008). A Genetic Algorithm Based Wrapper Feature Selection Method For Classification of Hyper spectral Images Using Support Vector. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 37. Part B7. Beijing, pp 397-402.
- Vidyavathi, B.M., Ravikumar, C. N. (2008). A Novel Hybrid Filter Feature Selection Method for Data Mining. *Ubiquitous Computing and Communication Journal*. Vol. 3, Issue 3, pp 118-121.
- Raymer, M.L. Punch, W.F., Goodman, E.D., Kuhn, L.A., Jain, A.K. (2000). Dimensionality Reduction using Genetic Algorithms. *IEEE Transactions on Evolutionary Computation*, Vol. 4, Issue 2, pp 164-171.
- Agarwal, Mayank, Jain Nikunj, Kumar Manish, Agrawal Himanshu. (2010). Face Recognition Using Eigen Faces and Artificial Neural Network. *International Journal of Computer Theory and Engineering*. Vol. 2. Issue 4. pp 624-629.
- Imani, Maryam Bahojb, Pourhabibi Tahereh, Keyvanpour Mohammad Reza, and Azmi Reza. (2012). A New Feature Selection Method Based on Ant Colony and Genetic Algorithm on Persian Font Recognition. *International Journal of Machine Learning and Computing*. Vol. 2, Issue 2, pp 278-282.
- Pali, Vivek and Bhaiya, Lalit Kumar. (2013). Genetic Algorithm Based Feature Selection and BPNN Based Classification for Face recognition. *International Journal of Advanced Research in Computer Science and Software Engineering*. Vol 3, Issue 5, pp 269-273. ISSN. 2277 128X. Available online at. www.ijarcsse.com. Accessed on 9th March, 2016.
- Hatua, Amartya. (2014). Optimal Feature Selection from VMware ESXi5.1 Feature Set. *International Journal of Chaos, Control, Modelling and Simulation (IJCCMS)* Vol. 3, Issue 3, pp 1-8.
- Fagbola, Temitayo., Olabiyisi, Stephen, Adigun Abimbola. (2012). Hybrid GA-SVM for Efficient Feature Selection in E-mail Classification. *Computer Engineering and Intelligent Systems*. Vol. 3, Issue 3, pp 17-28.
- Kashef, S., Nezamabadi-pour, H. (2014). An advanced ACO algorithm for feature subset selection. *Neurocomputing*. Pp 1-9. <http://dx.doi.org/10.1016/j.neucom.2014.06.067>
- Mauricio, Schiezero and Helio, Pedrini. (2013): Data feature selection based on Artificial Bee Colony algorithm. *EURASIP. Journal on Image and Video Processing*, 2013: 47. <http://jivp.eurasipjournals.com/content/2013/1/47> Accessed on 9th March, 2016.
- Dorigo, M, Birattari M, Stutzle T. (2006). Ant colony optimization. Artificial ants as a computational intelligence technique. *IEEE*

- Computational Intelligence Magazine*, Vol. 1. Issue 4. pp 28-39.
- Mehdi, Hosseinzadeh Aghdam, Nasser Ghasem-Aghaee, Mohammad Ehsan Basiri. (2009) Text feature selection using ant colony optimization. Elsevier *Journal of Expert Systems with Applications* Vol 36, pp. 6843–6853.
- Stutzle, Thomas, Lopez-Ibanez, Manuel, Pellegrini, Paola, Maur Michael, Montes deOca Marco, Birattari Mauro, and Dorigo Marco. (2010). Parameter Adaptation in Ant Colony Optimization. *IRIDIA Technical Report Series*. Technical Report No.TR/IRIDIA/2010-002.
- Jensen, R and Shen Q (2005) Fuzzy-Rough Data Reduction with Ant Colony Optimization. *Fuzzy Sets and Systems*, Vol 149. Issue 1, pp 5-20.
- Chen, Y. (2011). Ant colony optimization approach for stacking configurations (Master's thesis, Lingnan University, Hong Kong). Retrieved from http://dx.doi.org/10.14793/cds_etd.2. Accessed on 9th March, 2016.

Reengineering open source software for extended functionality: the case of Open Medical Record System (OpenMRS)

Iwara I. Arikpo^{1,2} and Anthony T. Okoro³

Department of Computer Science
University of Calabar, Nigeria

¹iarikpo@gmail.com, ²iwara.arikpo@unical.edu.ng and ³anthonyokoro1@gmail.com

ABSTRACT

Purpose: To proffer solution to the numerous problems of paper-based medical record systems predominant in low- and medium-income countries (LMICs), by reengineering an open source electronic medical record software system, the OpenMRS, for customized and extended functionality in hospitals.

Methodology: Software reuse-oriented model and reengineering was applied to extend the functionalities of the OpenMRS. Qualitative interviews were conducted to gather information, analysis of the OpenMRS' core architecture, bundled modules and form creation features were done to identify functional and non-functional requirements. The software development tools: Java, Maven, HTML, JSP, Tomcat, and MySQL were used for the reengineering.

Findings/Results: The study found major constraints in the management of hospital records and service delivery due to paper-based methods. An open source client-server electronic medical record system was successfully extended to accommodate features like efficient staff-patient scheduling, and facility mortality data, all at less cost and with minimal programming effort.

Research limitations: Inability to include all units of the hospital, such as pharmacy, accounting/billing, and others in the implementation, due to time and some technical hitches.

Practical implications: The study showed that, open source electronic medical record systems can be reengineered for extended and customized functionalities for hospitals in LMICs, with the overarching target of improving service delivery in these climes. The integration of facility-based and community-based software systems is recommended for further research.

Originality: Lots of research has been done and many still on-going in the fields of open source software and reengineering, but this study is the first to adopt an approach of a practical case study

of the reengineering of an open source medical record system.

Keywords: *Open source, reengineering, OpenMRS, electronic medical record system, software reuse, paper-based system.*

Paper Type: *Research Paper*

1. Background

Provision of medical services to patients is vital in every society, the success of which is dependent on the timely availability of patients' medical histories and other related records such as patients' demographic information. Efforts by hospital managements and other key players in the health sector to keep these records complete, up-to-date, and handy have seen hospital use paper-based methods – forms, registers, log books, etc. for the management of these critical patients and hospital information. This huge reliance on paper-based record keeping is termed “paper-based medical records system”. It is the type of system that depends majorly on the use of paper for collection and storing of hospital information with minimal or non-use of electronic systems.

Adoption of this poor records tracking and review system often results in delayed service delivery with attendant negative implications. For instance, Nigeria's maternal and child health clinics is still burdened by various paper-based health information management system problems despite the support by local and foreign agencies (Thompson et al., 2009). The application of Information Technology (IT) in health can help reduce this burden, minimize the cost of running hospitals and introduce the potential for sharing patients' data with fear of privacy (Idowu et al., 2008).

The overwhelming need for basic clinical data management along with the need for rapid response in the face of limited technical resources led to the development of many proprietary software products with no elaborate standards to guide use, and facilitate collaboration with other vendors to ensure compatibility (Goulde et al.,

¹ This paper is an extract from an MSc Work under the supervision of the lead author.

2006). The solution rests in the use of electronic medical record systems (EMRs).

EMRs provide benefits such as: enhanced information integrity and security, reduced transcription errors, reduced duplication of data entries and optimized report turnaround times. Other advantages of MRS are: Timely and controlled (authorized) access to patient's information – demographics, diagnosis, allergies, laboratory test results, previous prescriptions, patient's medical history and visits; result management, decision support, order management, internal and external communication between healthcare providers, patient's support, financial and other administrative processes such as stock management, staff scheduling and reporting (Tang, 2003). Unfortunately, the availability of a cost-effective system that integrates these services to meet the specific requirements of hospitals in most cases, is not readily in place. The concern is therefore, on the availability of an extensible turnkey system that is capable of managing hospital health information effectively at reduced costs, without much technicality required for its implementation. Open source systems can be relied upon to provide this anticipated solution.

"Open source" is a concept that denotes the class of software in which the source code is freely available for intending contributors or code reviewers. The use of these freely available program source codes either for the development of new systems or modification of the existing ones led to the software design paradigm known as open source development. It is usually a collaborative effort where software developers in different locations of the world access the source codes and make needed contribution with the overall goal of improving the quality of the software.

The overall goal of this study is therefore, to provide capacity for the adaptation and customization of an open source software. The implementation will eliminate the huge financial burden accruing from the outsourcing of EMRs development, reduce the rate of transcription errors arising from the use of paper-based tools, save cost on paper and storage space, improve staff performance, reduce patients waiting time at hospitals and extend the EMR functionalities to support facility-based mortality data collection. Subsequent sections will address health management information system solutions used in the past, the advantages and associated problems. It shall further present the open source solutions proposed in this work with the methods suggested for its easy adaptation and implementation.

1.1 Problems with paper-based and proprietary systems

Generally speaking, data is a critical item needed by health policy-makers to measure the impact of policies and interventions, assess progress of healthcare programmes and for proper planning. Sadly, the paucity of these health data and, in cases where it is collected, the delay in its availability restrain hospitals and healthcare providers from carrying out prompt healthcare services for patients in dire need of such services. Despite the use of traditional paper-based systems and proprietary software for collection and management of these data, hospitals still struggle to account for key health indicators. For instance, the near-absence of functional vital registration systems in LMICs affects the ability of health facilities to adequately account for deaths and causes of deaths in the hospitals.

As identified by Abdulkadir et al. (2011), problems associated with paper-based patient record-keeping in hospitals include: non-documentation of patients' complete demographic and clinical information, mutilations of medical records, illegible scribbles and frequent use of unclear abbreviations in many health facilities. In addition, there is high tendency of information loss as a result of paper damage and misplacement. Other setbacks concomitant with most health institutions are the issues of tracking staff work flow using paper-based staff rosters and movement books. The dynamic nature of hospital settings makes this approach difficult to maintain, track and adhere to.

Similarly, in situations where electronic systems are used, the cost and technical considerations involved in redesigning clinical processes and workflow to incorporate complex laboratory, clinic, finance and administrative structures of the health institution coupled with the cost of end-user training and support (Agency for Healthcare Research and Quality, 2015; Menachemi & Collum, 2011) often limits the efforts of the implementers. Attempts to use commercial-off-the-shelf (COTS) applications or paper-based medical records system does not proffer expected solution as these systems are not scalable and are difficult to maintain (Iyer, 2013; Oberndorf, 1998).

All these have negatively affected good management of patients' medical records and hospital information. The above challenges can be addressed if a feasible and cost-effective medical records management solution is well implemented using free, customizable and open source

applications like the Open Medical Record System (OpenMRS).

2. Literature Review

2.1 Developments in Medical Records Systems

A medical record system is a crucial component in every modern day health institution. It provides facilities for record-keeping of patients' medical information. Interestingly, before the birth of electronic medical record keeping systems, health practitioners only diagnose the cause of sickness in a patient by mere looking at the patient's complexion, pulse or urine without any consideration on the need to document the patient medical information after such diagnosis and treatment (Casebook Project, 2016). The practice continued till about the seventeenth century when Astrologers, Forman and Napier produced a casebook for recording list of clients, their payments and prescription. The casebooks were systematically produced, and provided the best sources of information reference between patients and the medical practitioners (Kassell, 2014). Gradually, the use of books for recording accounts information and medical histories, followed. Medical practitioners began documenting the observations during examination of patients and then used the results to advice on better and safer lifestyle practice. With the steady transformation of the world into information age in the 1920s, the paper medical records system sprang up.

In 1928, the United States officially established the Health Information Management (HIM) System to enable the American College of Surgeons (ACOS) improve the standard of data collected in clinical settings (Brooks, 2015). The benefits of documenting patient data were obvious as it was possible to easily see patients' background information, complications and outcome of care. The method of medical record keeping was paper-based until the 1960s and 70's when the emergence of computers stimulated researchers on the possibility of building medical records into computers (Brooks, 2015). In 1972, the first medical records system was developed by Regenstreif Institute even though it did not receive wide acceptance by many physicians. The information age revolutionized the health management information system. Systems were developed to record the data of a patient electronically at a particular facility. As at 1980s, the trend was the development of single computer application by vendors for specific departments. The disadvantage was that the system could only

serve the purpose for which it was built; new developments requiring modification of requirements and functionalities were difficult to implement.

2.2 Introduction of Electronic Medical Record Systems

As highlighted by HealthIT.gov (2015), Electronic Medical Record (EMR) systems "are digital versions of the paper charts in clinicians' offices, clinics, and hospitals. EMRs contain notes and information collected by and for the clinicians in that office, clinic, or hospital and are mostly used by providers for diagnosis and treatment." An electronic health record system is a system that provides each individual with a secure and private lifetime record of their key health history and care within a health system. The record is available electronically to authorized healthcare providers and the individual anywhere, anytime, in support of high quality care (Neville et al., 2004). These systems have gradually transformed from the complete manual processes of health data collection, cleaning, storage and retrieval to a more stable and reliable information technology-driven systems with all clinical, financial and administrative applications fully integrated (OpenClinical, 2013). As shown in Figure 1, an EMR provides a centralized repository that is capable of storing patients' health data (from various units/departments) efficiently. These data can also be accessed by different people at the same time (Krishna, 2010).

Though the emergence of electronic medical record systems have reduced the difficulty inherent in traditional paper-based systems, there are still limitations to its global use, especially with the customer-ever changing demands and requirements. As was the practice, most of the developed systems were proprietary, while others were tailor-made. The proprietary systems are finished products usually developed with no user in mind. They serve as a general purpose systems. For the tailor-made or bespoke products, developers work on the software based on requirements of a particular user.

Whichever approach is used for the development of these electronic systems, users' implementation experience vary. Some users record successful implementation while others failed. For instance, Cedars-Sinai Medical Center, Los Angeles in 2002 lost over US\$34 million in attempt to implement a new electronic medical record system.

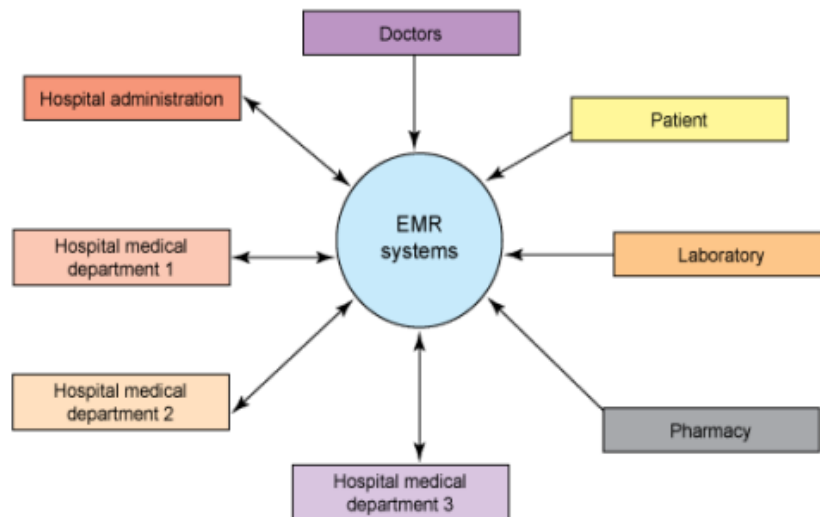


Figure 1: A simple electronic medical system (Krishna, 2010)

The major reason for the failure was huge time wastage by physicians in using the order entry system for medications, lab, and other treatment procedures. Unnecessary time was spent in responding or avoiding system notifications and warnings (Kumer & Aldrich, 2010). Unfortunately, despite the huge finance and time spent on the project, its implementation failed.

Another challenge preventing the wide use of some robust electronic medical record systems is the cost of implementation and maintenance. For example, VisTA (the Veterans Health Information Systems and Technology Architecture) is one of the precursors to modern Electronic Health Record (EHR) programs used mostly in the United States. It still presents installation and maintenance constraints. Consultants are usually hired for its configuration; a process which requires huge service charge. It is also not user-friendly (Medved, 2014). So, even though these systems are reliable, the cost of implementation especially in resource-poor settings, inability to reengineer and modify or add new functionalities, etc. still pose challenges.

For tailor-made systems, the functions are always limited to its specific requirements and at such not suitable for other units. ZEPRS (Zambia Electronic Perinatal Record System) is a typical example of such EMR. It is used by public obstetric clinics in Zambia to improve patients care and manage the records of only perinatal patients and infant clinics (RTI International, 2015). The software's limited use in Zambia remains a major doubt of its robustness in other regions and health care settings.

The quest to improve electronic medical record keeping continued with the introduction of open source technology. The paradigm provided capacities to customize the system to suit the needs of the user. This flexibility is evident in the numerous open source systems in use today. Examples include DHIS2, OpenEHR, OpenEMR, BlueEHS, GNUmed and OpenMRS.

OpenMRS is web-based patient medical record system developed by teams of medical and health experts and has proven to be a reliable and robust system for management of medical records. It also provides a customizable interface and extendable modules which are free and readily available. The software is presently implemented in over 50 countries, including the US, South Africa, Germany, the Philippines and Chile. Further, both the Rwandan and Kenyan governments have chosen the software for the implementation of their national e-health infrastructure (Connelly, 2012). With access to free and open source software and the application of computer science methodologies such as software reengineering and reuse, the cost of implementing electronic medical record systems in resource-poor settings can be minimized with improved quality of healthcare service delivery.

2.3 Software Reengineering and Reuse

As defined by Chikofsky and Cross (1990), "reengineering is the examination, analysis and alteration of an existing software system to reconstitute it into a new form, and the subsequent implementation of the new form" (p.15). To reengineer a software will entail recovering and reusing existing software assets (products and by-

products of the software development life cycle and include code, software components, test suites, designs and documentation, etc.), putting high software maintenance costs under control, and establishing a base for future software evolution (Rosenberg, 2015).

Reengineering is a technique that has shown to improve existing systems for easy understanding, control and use. Apart from reducing the huge dependence on personnel efforts, it increases maintainability, interoperability, performance, and testability of existing systems (Abbas et al., 2012). Many legacy systems have been structured and modularized through reengineering to yield better systems. This study was designed to achieve the objective of applying reengineering concepts and software reuse on an open source medical record system, the OpenMRS.

3. Methodology

Software reuse-oriented model (Figure 2) and reengineering technique (Figure 3) were used in this study to extend the functionalities of an open source electronic medical record system to provide a centralized data repository, manage staff-patient work flows and capture all data on

facility-based deaths. The study, conducted between October 2015 and March 2016 began with requirement analysis, interactions with stakeholders (physicians, medical students, patients, and record officers) in the University of Calabar Teaching Hospital (UCTH). This requirement gathering phase assisted in the identification and refinement of the functional and non-functional requirements. Technical and best practice guidelines for the installation and setup of the EMR system, build of additional concepts and features into the system were gotten from the online resource kits, OpenMRS wiki pages, medical experts and hospital records management staff.

The second phase of software reuse allowed for the analysis of the core and bundled modules of the software. Functional requirement issues like the report generation, appointment scheduling, patient registration, encounters, observations, creation of concepts, in addition to considerations on the extendibility of the core application to accommodate new modules were identified. Maven project management (Maven, 2016) tool facilitated the reengineering of the application to enable customization of the interface and form entry module.

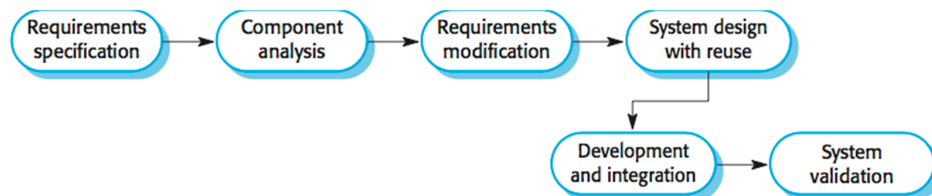


Figure 2: Reuse-oriented model (Sommerville, 2011, p. 35)

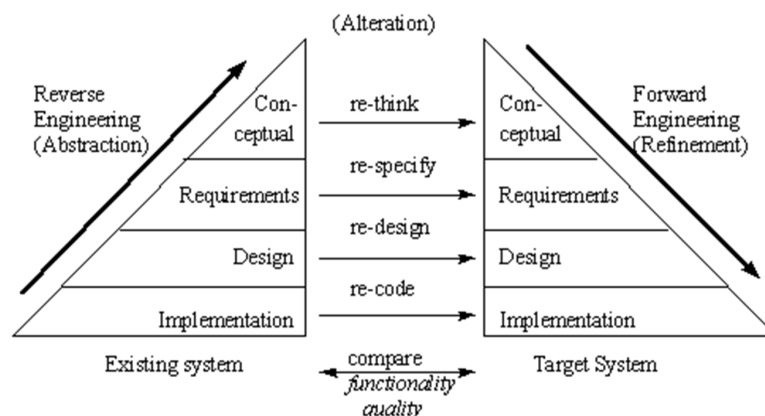


Figure 3: General model for software reengineering (Byrne, 1992)

In performing the build process, the source code was forked from the online repository into Eclipse Integrated Development Environment (IDE) for

code modification (Eclipse, 2016). Maven was then used to build the modified codes; generating a 'war' (Web Application Archive) file and .omod

files for deployment on the Tomcat server. A 'war' is a JAR file used to distribute collection of JSP, Java Servlets, Java classes, XML, etc. It allows for easy testing, deployment and identification of the version of deployed application. On the other hand, .omod (an extension for OpenMRS module) is the package extension format created with maven for the modules. It is a renamed JAR file containing the config.xml file (this file holds basic information about the module).

A tool for the collection of facility-based deaths was created and embedded into the HTML form entry module. This was done using the specialized OpenMRS-HTML tags. The data collection tool was designed to store every observation as concept in the MySQL database. Hibernate mapping tool (Hibernate, 2016) provided the mapping of the objects in the data model to the tables in the database.

In addressing the requirement for patient-staff scheduling, an appointment module was deployed and configured into the system with minimal programming effort, except for technical skills in identification and configuration of dependent modules. The overall system was deployed and tested on a Dell PowerEdge server running Windows 2008 Server R2 operating

system in the Cross River Health and Demographic Surveillance System (CRHDSS), University of Calabar, Nigeria. The local area network infrastructure provided access for user login and testing.

4. Results

With open source applications, users are at liberty to restructure systems to their needs. This was demonstrated in this work through the customization of the form entry module and user interface, as shown in Figure 4. One of the strengths of open source software in general, and the system under study in particular, is the high level of extendibility to accommodate unanticipated functionalities. This is demonstrated in this study through the creation of additional tool for the collection of all facility deaths, not part of the original OpenMRS adapted. The system works by identifying all deceased patients as 'deceased' (Figure 5). The extended tool is then invoked as an encounter where the death information (symptoms leading to death, diagnosis, date of death, etc.) are entered. Upon creation of this encounter, the system generates a summary of the death encounter as shown in Figure 6.

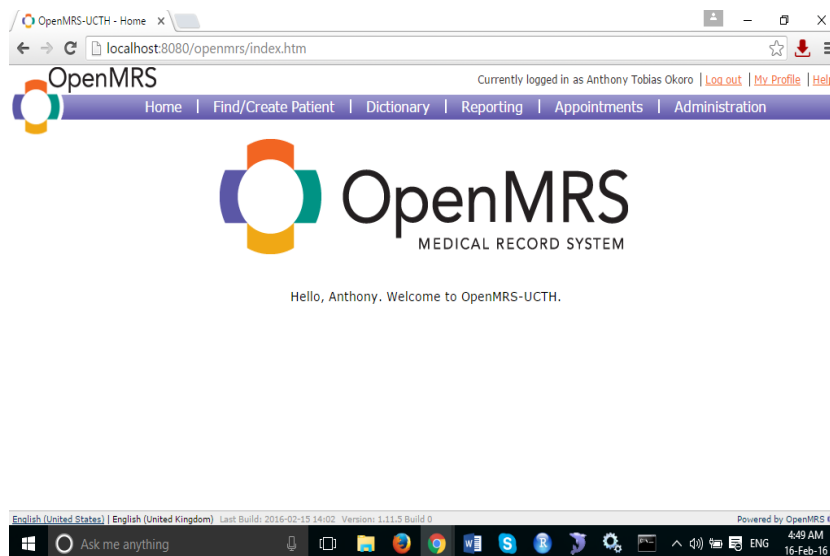


Figure 4: Customized Home Screen

OpenMRS-UCTH - Patient X

localhost:8080/openmrs/patientDashboard.form?patientId=18

OpenMRS

Currently logged in as Anthony Tobias Okoro | [Log out](#) | [My Profile](#) | [Help](#)

Home | Find/Create Patient | Dictionary | Reporting | Appointments | Administration

This patient is deceased Death Date: 10/02/2016 Cause of Death: OTHER NON-CODED : typhoid

Imoh morad Hosp No: em321234

56 yrs (08-Feb-1960)

BMI: ? (Weight: , Height:) CD4: | Regimen:

Active Visit: [Consultation Visit](#) at emergency from 12/02/2016 [Edit Visit](#) [End Visit](#)

No encounters yet

Active Visit: [Consultation Visit](#) at ENT from 01/12/2015 [Edit Visit](#) [End Visit](#)

No encounters yet

Overview Visits Demographics Graphs Form Entry Appointments

Last 3 Encounters

View	Encounter Date	Encounter Type	Visit	Provider	Form	Location	Enterer
View	14/02/2016	FacilityDeath		Emmanuel Gibson	Mortality Records Surveillance	UCTH	Anthony Tobias Okoro

Enter Form

Search: ☐ Include Retired

Figure 5: A patient marked as dead in the system

2/16/2016 OpenMRS-UCTH

OpenMRS

Currently logged in as Anthony Tobias Okoro | [Log out](#) | [My Profile](#) | [Help](#)

Home Find/Create Patient Dictionary Reporting Appointments Administration

[Back](#) | [Print](#) **Cosmos Agba | Mortality Records Surveillance (FacilityDeath) | 16/02/2016 |** [Edit](#) | [Delete](#)

Mortality Records Surveillance (v0.1.1) Paper Form ID: (MS01)

1. Encounter Details

Date: 16/Feb/2016
Location: Casualty Ward
Provider: D

2. Demographic Information of Deceased

Name:
Hosp No: ca568376
Gender: M
Birthdate: ~ 01/Jan/1914 (Age: 102)
Citizenship: nigeria
Senatorial District: southern
LGA: akamkpa
Community: akamkpa
State: Cross River
Health Facility: UCTH

3. Deceased Area of Residence Identification Data

A0: State Cross River A1: Senatorial District: southern
A2: LGA: akamkpa
Pregnant or recently delivered (with past 6 weeks): [] Yes [X] No
Last Date of Arrival at Facility: 31/May/2014 Time of Arrival at Facility (hh:mm:ss): 11:20
Date of Death: 02/Jun/2014 Time of Death (hh:mm:ss): 11:24

4. Symptoms and Diagnosis

☐ Cough	☐ Fever	☐ Vomiting	☐ Diarrhea	☐ Constipation	☐ Breathing Difficulty	☐ Convulsion
☒ Unconscious	☐ Jaundice	☐ Bleeding	☐ Very Pale	☒ Accident (RTA)	☐ Growth/Swelling	☐ Unknown
						Other (specify) none

5. Diagnosis / Likely cause of death

☐ Malaria ☐ Typhoid ☐ Pneumonia ☐ Dysentery/Diarrhea ☐ Severe Acute Malnutrition ☐ Other (specify) severe head injury
☐ Diabetes Mellitus ☐ Stroke ☐ Hypertension ☐ Child birth ☐ Burns ☐ HIV/AIDS ☐ Tuberculosis ☐ severe head injury
☐ (TB) ☐ Gastro-enteritis ☐ Birth Asphyxia ☐ Tetanus ☐ Cancer ☐ Injuries ☐ Unknown

6. Second Diagnosis / Additional likely cause of death

☐ Malaria ☐ Typhoid ☐ Pneumonia ☐ Dysentery/Diarrhea ☐ Severe Acute Malnutrition ☐ Other (specify) severe head injury
☐ Diabetes Mellitus ☐ Stroke ☐ Hypertension ☐ Child birth ☐ Burns ☐ HIV/AIDS ☐ Tuberculosis ☐ severe head injury
☐ (TB) ☐ Gastro-enteritis ☐ Birth Asphyxia ☐ Tetanus ☐ Cancer ☐ Injuries ☐ Unknown

English (United States) | English (United Kingdom) Last Build: 2016-02-15 14:02 Version: 1.11.5 Build 0
http://localhost:8080/openmrs/module/htmlFormEntry.html?formEntryForm?encounterId=10

Powered by OpenMRS 1/1

Figure 6: Summary of facility death encounter of a deceased patient created in the reengineered OpenMRS

The problem of patients' delay was also addressed in this study, by deploying and configuring a staff-patient scheduling feature in such a way that every provider, on daily basis, knows the patients coming for medical care. The scheduling indicates the type of visit; which could be return visit, preventive return care, routine problem visit, follow-up visit, etc. Patients' long wait at hospitals is overcome, since prior to the visit, the physician is already aware of the patient's potential visit. This enables the health care provider to review the case file (medical history) of the patient. This can also aid in the reduction of staff redundancy since it guides the administrators in checkmating the daily activities of staff. Besides, is also easy to generate quick summaries from the patient data stored in the repository. This implementation presented in this study provided facility for generating reports on individual demographics, programs enrolled, symptoms, diagnosis, etc.

5. Discussion

Adaptation of software products, the codes of which are free and open, can reduce users' anxieties from the use of proprietary systems. Software development methods such as reuse and reengineering, are efficient computer science approaches equipped for the implementation of open source solutions at minimal cost.

As demonstrated in this study, open source solutions can be tweaked to provide additional functionalities, based on the ever changing requirements of users over time. These modifications can be achieved at very low cost and with minimal technical expertise. The rich documentation available for each open source solution can guide an implementer who has intermediate programming and computer science skills to deploy and use the solutions correctly. For instance, the facility-based mortality data collection tool is a feature added to the OpenMRS by applying the knowledge of Java, JSP, HTML scripting and the Maven project building tool. This was made possible through the software's documentations, technical guidelines on embedding data collection forms into the core application; such that the concepts and encounters are linked and overall structure maintained.

Software reuse closely assists in the achievement of a new and improved software product from an existing open source solutions. There are usually a myriad of components that have been developed and tested by experts for a given functionality. These components are freely available in online repositories, though some may

attract minimal cost for reuse. The only tedious tasks usually faced here is the ability of the programmer/developer to identify the component that best fits the required functionality. Having identified the necessity of efficient staff-patient scheduling and reporting, this study was able to adapt and modify existing scheduling and reporting module of the OpenMRS into the core application. With minimal programming effort and technical skills, the deployed module saved the cost of designing and developing a staff-patient scheduling system from scratch. Integration and interoperability concerns often associated with cross-platform applications was overcome through this implementation. This is, however, not the case with proprietary and non-open source systems.

6. Conclusion

The study explored the possibility of extending the functionalities of open source software products through software reuse and reengineering techniques. With the use of software reuse method, bundled modules for staff-patient appointment scheduling and reporting were identified, analyzed, configured and deployed into the core application. Similarly, the reengineering technique helped in the alteration of the core application to permit addition of a data collection tool. A template for recording all facility-based deaths was designed and embedded into the system. Customization of the user interface and HTML data entry module was also implemented.

Successful implementation of this reengineered application, the OpenMRS, can help strengthen the data management capabilities in health facilities. The burden often faced by health professionals, research experts and policy makers in having complete and reliable data for timely formulation of policies and planning health care interventions will be drastically reduced. In terms of contribution to knowledge, this paper has demonstrated the elasticity of open source software in a tertiary hospital setting; an approach that can be extended to other hospitals in LMICs. The limitation of this study, is the inability to include all units of the hospital, such as, pharmacy, accounting/billing, etc. in the implementation, due to time and some technical hitches.

6.1 Suggestion for future work

Integrating community-based systems like the Open Health Demographic System (OpenHDS)

with facility-based medical record systems such as OpenMRS can help policy makers have complete data for comparative analysis of disease burdens and measure impact of health interventions.

References

- Abbas, S., Jeberson, W., & Klinsega, V. V. (2012). The Need of Re-engineering in Software Engineering. *International Journal of Engineering and Technology*, Vol 2, p 2.
- Abdulkadir, A., Yunusa, G., Tabari, A., Anas, I., Ojo, J., Akinlade, B., Suleman, B., Uyobong, I. (2011). Medical record system in Nigeria: observations from multicentre auditing of radiographic requests and patients' information documentation practices. *Journal of Medicine and Medical Science* Vol. 2, No 5 pp 854-858.
- Agency for Healthcare Research and Quality. (2015). Electronic Medical Record Systems. Retrieved February 2016, from <http://healthit.ahrq.gov/key-topics/electronic-medical-record-systems>
- Brooks, A. (2015). Health Information Management History: Past, Present & Future. Rasmussen College. Retrieved February 2016, from www.rasmussen.edu/degrees/health-sciences/blog/health-information-management-history.
- Byrne, E. (1992). A conceptual foundation for software re-engineering. Proceedings for the Conference on Software Maintenance, Orlando, Fla., 226-235. <http://dx.doi.org/10.1109/ICSM.1992.242539>
- Casebooks Project (2016). History of medical record-keeping. Retrieved June 2016, from <http://www.magicandmedicine.hps.cam.ac.uk/on-astrological-medicine/further-reading/history-of-medical-record-keeping>.
- Chikofsky, E. & Cross, J. (1990). Reverse Engineering and Design Recovery: A Taxonomy. *IEEE Software*, Vol 7, No 1, pp 13-17. <http://dx.doi.org/10.1109/52.43044>
- Connelly, M. (2012). OpenMRS implementers: Experiences and Lessons Learned. *Global Health Delivery Project at Harvard University*. <http://www.globalhealthdelivery.org/blog/2012/05/openmrs-implementers/> [Retrieved:]
- Eclipse (2016). *Eclipse IDE for Java EE Developers*. Retrieved 18 May 2016, from <http://www.eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/marsr>.
- Flossmanuals. (2015a). OpenMRS Developers Guide. En.flossmanuals.net. Retrieved from http://en.flossmanuals.net/_booki/openmrs-developers-guide/openmrs-developers-guide.pdf
- Goulde, M., Brown, E., Rymer, J., Bough, V., & Hartzband, D. (2006). Open Source Software: A Primer for Health Care Leaders. iHealth Reports. *California HealthCare Foundation*.
- Groen, P. (2013). OpenMRS Installations & Users in Africa. Open Health News. Retrieved April 2016, from <http://www.openhealthnews.com/hotnews/openmrs-installations-users-africa>
- HealthIT.gov. (2015). What are the differences between electronic medical records, electronic health records, and personal health records? Retrieved July 2016, from <https://www.healthit.gov/providers-professionals/faqs/what-are-differences-between-electronic-medical-records-electronic>.
- Hibernate (2016). Retrieved 18 May 2016, from *Hibernate Tools*. <http://hibernate.org/tools/>.
- Idowu, P., Cornford, D., & Bastin, L. (2008). Health informatics deployment in Nigeria. *Journal of Health Informatics in Developing Countries*, Vol 2, No 1, pp 15-23.
- Iyer P. (2013). 5 Top Disadvantages of Manual Medical Records. Retrieved March 2016, from <http://www.medleague.com/blog/2013/04/01/5-top-disadvantages-of-manual-medical-records/>.
- Kassell, L. (2014). Casebooks in Early Modern England: Astrology, Medicine and Written Records. *Bulletin of the History of Medicine*, Vol 88, No 4, pp 595-625.
- Krishna, S. (2010). Taking medical records into the digital age. *IBM*. Retrieved March 2016, from <http://www.ibm.com/developerworks/webstore/library/techarticles/indopenemr/#fig1>.
- Kumer, S. & Aldrich, K. (2010). Overcoming barriers to electronic record (EMR) implementation in the US healthcare system: A comparative study. doi: 10.1177/1460458210380523
- Maven (2016). *Apache Maven Project*. <https://maven.apache.org/> [Retrieved: 18 May 2016].
- Medved JP. (2014). The Top 7 Free and Open Source EMR software Products: Capterra Medical Software Blog. Retrieved July 2016, from <http://blog.capterra.com/top-7-free-open-source-emr-software-products..>

- Menachemi, N., & Collum, T. (2011). Benefits and draw backs of electronic health record systems. *Risk Management and Health Policy. Dove Press Journal*, Vol 4, pp 47-55.
- Neville, D., Gates, K., Tucker, S., Keough, M., MacDonald, D., Barron, M., O'Reilly, S. (2004). Towards an Evaluation Framework for Electronic Health Records Initiatives: An Annotated Bibliography and Systematic Assessment of the Published Literature and Program Reports. Retrieved July 2016, from http://www.nlchi.nf.ca/pdf/bio_feb04.pdf.
- Oberndorf, P. (1998). *Software Engineering Institute Monographs on the Use of Commercial Software in Government Systems. COTS and Open Systems*. Retrieved March 2016, from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.5124&rep=rep1&type=pdf.
- OpenClinical. (2013). Electronic Medical records, Electronic Health Records. Retrieved March 2016, from <http://www.openclinical.org/emr.html>
- Rosenberg, L. (2015). *Software Re-engineering*. Retrieved February 2016, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.447&rep=rep1&type=pdf>.
- RTI International. (2015). ICT Improves Patient Care in Africa: ZEPRS -- An Electronic Perinatal Record System in Zambia. Retrieved February 2016, from <http://www.rti.org/page.cfm?objectid=09C970A2-AB6E-504A-37FCAF624F0E9927>.
- Sommerville, I. (2011). *Software Engineering* (9th Ed.). Pearson Education Inc., USA.
- Tang P. (2003). Key Capabilities of an Electronic Health Record System. Institute of Medicine Committee on Data Standards for Patient Safety. Board on Health Care Services. Washington D.C.: *National Academies Press*. Retrieved February 2016, from <http://www.nap.edu/catalog/10781.html>.
- Thompson A., Castle E., Lubeck, P. & Makarfi, S. (2009). Experience Implementing OpenMRS to Support Maternal and Reproductive Health in Northern Nigeria. Retrieved February 2016, from http://ehealthafrica.org/wp-content/uploads/2009/10/OMRS_MedInfo_2009_v1.pdf.

A Mobile-Based Phishing Detector App for E-Mails

Abdul A. Orunsolu*, Misturah A. Alaran, Sakiru O. Kareem, Ayobami G. Oke, Eniola I. Adejonwo

Department of Computer Science,
Moshood Abiola Polytechnic
P.M.B. 2240, Abeokuta, Ogun State, Nigeria
*orunsolu.abdul@mapoly.edu.ng

ABSTRACT

Purpose: The objective of this study is to design and implement a light weight mobile application for detecting fake e-mail messages on smart phones.

Design/Methodology/Approach: The design is based on the characteristics or features of Universal Resource Locator (URL) with the e-messages. These URL's features are then compared with existing "fingerprints" of malicious URLs in a rule-based approach.

Findings: The results indicated that the approach is very effective in detecting phishing attacks on mobile phones with a True Positive of 94.5% and True Negative of 100% on the collected dataset.

Research limitations/implications: The proposed technique is tested with datasets from known phishing and benign data sources. The technique can provide the needed security for mobile phones users.

Practical implications: This work provides the design of a mobile based phishing detector app which can verify the identity of embedded links within e-mail messages. The practicability of the proposed solution was tested using Gmail and Yahooemail as their Application Programming Interface (API) is easily available for integration

Originality/value: The popularity of mobile devices especially for online-related activities has made most unsuspecting users vulnerable to web-based attacks. The proposed technique provides a secure, lightweight and efficient solution for mobile phone users in conducting their online activities especially e-mail management.

Keywords: *E-mail, Mobile devices, Phishing, Security, URL*

Paper Type: Research Paper

1. Introduction

Over the past few decades, following the growth of communication networks, the use of internet has become widely popular, with millions of new users subscribing to the online communication. Almost all shades of human endeavor now offer

their products and services on the internet to reap the benefits of global coverage, automatic availability of service, instant acknowledgement, speed etc. However, these positive incentives are now being challenged more than before by the marauding con-artists who now use internet for perpetuating crimes. Today, financial crimes are transformed from direct attacks into indirect attacks with significant negative impact on the global economy (APWG, 2016). One of these indirect attacks on the internet that has caused huge financial losses is phishing attack. In 2015, the Nigerian Financial Sector lost about \$470 million to this attack. Similarly, an online computer security agency called RSA reported that over \$1.6 billion was lost in 2013 globally to phishing attacks (RSA, 2014).

Phishing is a cyber-security threat which uses social engineering techniques and technical subterfuge to trick internet users into revealing their confidential information. These pieces of confidential information are then used by the phishers to initiate transactions behind the victims (Chiew, Chang, Sze, & Tiong, 2015). These transactions cause damages ranging from financial losses to personal defamation. In the past, the threat of phishing is limited to users of PC and notebooks due to relative high cost of these devices.

However, today, the attack surface for phishing attacks is gradually shifted to mobile phones users where avoidance behavior is very low (Arun, 2016). Nowadays, touch screen smartphones have become dominant in the mobile phone market. However, typing on a virtual keyboard is not as easy as on a physical keyboard due to lower input accuracy, particularly when walking or sitting in a moving vehicle. Because of that, it is tempting to follow links in web pages or e-mails rather than typing the links manually. Another factor is that in smartphones, switching among applications or even shifting to other pages within a browser, is more complicated and tedious than when performed on a PC. Users who value convenience usually prefer to follow links from

other applications. In the light of this, phishing attacks succeed because users become accustomed to entering their credentials in familiar, repeated login interfaces (Shahriar, Klintic, & Clincy, 2015).

In this paper, we report a lightweight mobile phone-based anti-phishing detector application. The application is a Rule-based system where certain behavior of malicious URLs are used in the design process. The application is implemented as a playStore apps which can be synchronized with Gmail and YahooMail.

The rest of the paper is organized as follows: Section II discusses literature review and related work. In section III, the proposed architecture is presented and discussed. The experimental results and evaluations are discussed in section IV while section V concludes the work.

2. Literature Review

According to APWG Phishing Activity Trends Report (APWG, 2016), more than 75% of phishing attacks target retail services, online payment systems as well as financial institutions. Similarly, Aaron & Rasmussen (2015) revealed that more than 82% of phishing attacks target e-Commerce, banks as well as money transfer industries. To combat the high incidences of phishing attacks, different methods have been proposed in order to defeat phishing attacks. These approaches are classified into the followings:

- List based methods (e.g. Whitelist or blacklist)
- URL evaluation
- Visual assessment
- Heuristics techniques
- Machine Learning methods
- User awareness methods

Numerous tools are designed using whitelist or blacklist approach. For example, the Google Safe browsing is one of the popular anti-phishing implementations based on blacklist approach. Most of these tools authorize or reject a website through one or more references. The weakness of this approach is the lack of scalability and the ability to update the references quickly. Since creating websites often is cheap and easy, and their average life span may be a few hours, this approach in itself has low efficiency in preventing phishing attacks. On the other hand, evaluation of apparent characteristic of webpage or its web address, is one of the best and simplest methods

for phishing attacks detection, on which many studies have been conducted. In this approach, authors attempt to identify phishing attacks through analyzing website's information such as site logo, webpage visual similarity, security issues, and other webpage characteristics. Similarly, some approaches use the URLs characteristics of a link to determine the status of such site. These methods are generally referred to as heuristics approach. This approach was used by Tan, Chiew, & Sze, (2017) and other researchers.

The heuristic approach firstly collects a set of discriminative features (called heuristics) that can separate phishing activities from legitimate ones, then train a machine learning model to predict phishing attacks based on these features and finally use the model to recognize phishing attacks in the real world. In heuristics approaches, webpages or emails are treated as documents in which features can be extracted from the header, content and body. The machine learning techniques are used to increase the accuracy of phishing detection especially for zero-day attack in which a new phishing scam has just been launched. One of the most widely used ML algorithms is the Support Vector Machine (SVM) (Khonji et al. 2013).

Apart from software enhancement methods, the user awareness is designed to correct the gap left in the technical methods (i.e. software enhancement methods) because "no matter how intelligent a system is, the person behind the keyboard can make or mar the security of a system" (Hong, 2012).

This current work is premised on the popularity of mobile phones, which has given many cyber criminals an opportunity to expand the reach of their illicit business. Although there are numbers of anti-phishing solutions, only few are available for mobile phone users. Due to this, most mobile phone users fall victims of malicious codes (Shahriar, Klintic, & Clincy, 2015). Moreover, most smart phone users are not usually security conscious and therefore needs a solution that can protect them. (Orunsolu, Bamgboye, Alaran, Sodiya, & Omorinola, 2015). In addition, our search on the Google Play Store for e-mail anti-phishing apps returns a very poor result.

3. The Proposed System

The architecture of the proposed approach is presented in Figure 1. The objective of the system is to design an efficient anti-phishing email scheme using Rule-based method. The proposed system consists of three major components which are:

- Email Loader
- Detection Engine
- Alert Notification (Alerter)

3.1 Email Loader

The proposed solution contains a mini browser which displays the email of the user. These

components display email just like the mobile application of the email providers in order to enable the user have access to features they enjoy on their official email clients. The e-mail loader only supports two email providers which are:

- Yahoo Mail (Yahooemail)
- Google Mail (Gmail)

The interface for the email loader is presented in Figure 2. The user selects their email provider and provides their login details to have access to their messages. These messages can be read by the detection engine to determine their status.

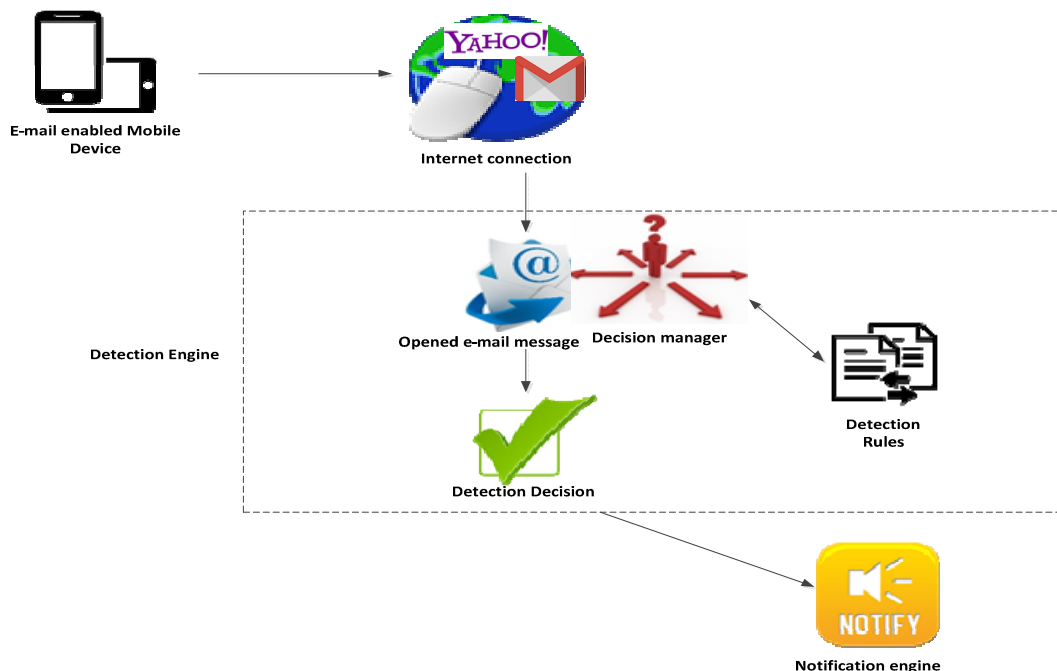


Figure 1: Architecture of the Propose Scheme

3.2 Detection Engine

This is a part of the application that is responsible for the scanning and detection of phishing sites. This component uses some set of rules to verify the authenticity of a URL in the user's inbox. Once a URL matches the criteria of a rule in the detection engine, the application then notifies the user about it.

The anti-phishing rules are:

1. URL ending with file name/extensions.
2. URL having more than 3 dots.

3. URL with special characters like @, #, \$, ~, -, ^, etc.
4. URL having IP addresses.
5. Website extensions appearing twice in URL

3.2.1 Filename/Extension Rule

Most of the phishing URL's in the Phishtank database end with certain file extensions like .php. However, legitimate sites usually end with web extensions like .com, .org as it is more secured than .php. Most .php extensions make a site easy

to attack. Some of the extensions noticed on most phishing URLs are:

- .html
- .htm
- .php

Examples:

<http://everydaycareer.com/igweabana/remade/login.htm>

<http://santamagda.cl/wp-content/languages/index.html>

<http://band.novahue.org/upCopy/gmail.php>

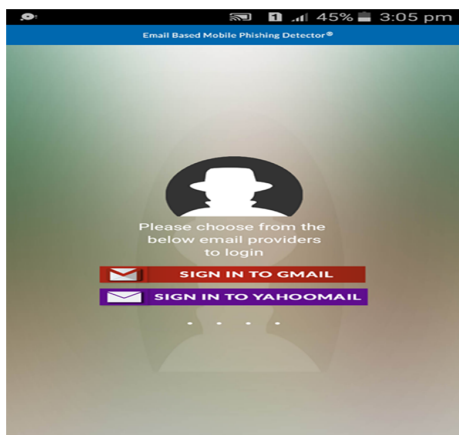


Figure 2. GUI of the email loader

3.2.2 URL having more than 3 dots

One way to mislead users to avoid detection of the primary webpage address is using high number of sub-domains in the URL path. In this way, attacker prevents users from identifying the actual website by creating multiple sub-domains and inclusion of related bank names in the following domains. For example in a page with the following URL, at first glance for a novice user, it may be assumed that the address is related to PayPal website but actually, this address is pointed to com.mx website:

<http://paypal.co.uk.webdllscr.mpp.sipromedic.com.mx>

In some previous studies that this feature is used, a specified number has been used as a way of detecting the suspicious addresses. For example, on average, in a URL address of a legitimate site there are maximally three dots. If the numbers of dots are more than this number, it will indicate that the webpage is phishing.

Example of a 3 dot website:

<http://www.google.co.uk>

Example of URL with more than 3 dots:

<http://vps38106.vs.webtropia.com/2016/login.php>

3.2.3 Using Special characters

This involves using special symbols in the URL of a website and can be used to redirect to a phishing website. Some of the URL's noticed are:

<http://www.gaya-maroc.ma/~verysign>

<http://gvmponda.com/css/alibaba/index.php?email=abuse@diplobel.fed.be%20&.rand=13vqcr8bp0gud&lc=1033&id=64855&mkt=en-us&cbcxt=mai&sns=1>

<http://conquerathletic.com.au/wp-includes/js/drop/index.php>

3.2.4 Using IP Address

One way to hide the webpage address is using an Internet Protocol (IP) address instead of a Domain Name System (DNS) address. Therefore, in this way the attacker prevents users to identify the actual page address. Another reason for using the IP address is that attackers often do not want to spend money to buy a domain for their fake webpages, and on the other hand, despite the possibility of tracking, they prefer to offer the host IP address to their users. Cyber criminals usually use this method to defraud unsuspected users in order to steal important information.

Examples:

<http://91.223.89.28/>

<http://gvmponda.com/css/alibaba/index.php?email=abuse@diplobel.fed.be%20&.rand=13vqcr8bp0gud&lc=1033&id=64855&mkt=en-us&cbcxt=mai&sns=1>

3.2.5 Website Extension Appearing Twice

It is also a common practice for phishers to use repetition of website extension in tricking unsuspecting users to navigate to fraudulent sites.

Example:

<https://googledrive.com/host/0Bxg87auMwzLATjRwdVlJekFXZWM?google.com>

However, the number of rules for our app are few to make it lightweight and we equally observed that some phishing rules have changed due to certain practices by web creators.

3.3 Rules That Have Changed

The rules presented in this section have changed considerably over time. These changes now make these URLs characteristics to be common to both phishing and legitimate webpages. Due to this observation, especially from our analysis of PhishTank and Alexa database, we have decided not to include them in our rules to reduce false positives and negatives. The URL characteristics are presented in the rest of this section.

3.3.1 Multiple Slash Rule

The multiple slash rule involves using more than 3 slashes “/” in the URL and this might be used to detect illegitimate websites.

Examples:

Phishing URL:

http://aandgtrading.com/category_images/sim/bookmark/ii.php?amp
<http://floridagunsupply.com/skin/install/pdf/home/service/costumer/information/check/fb5f7dfb371bb587e16de1124ada2d08/index/web/58b12f9a8c4b1ac5b9639243473c2560/login.php>

Legitimate URL:

<https://mail.google.com/mail/u/0/h/kd8qtu71x1sm/?zy=g&f=1>

Note: Looking at the URL of the legitimate sites and the phish URL's retrieved from Phishtank, it is observed that the number of slashes “/” exceeds three. This shows that we cannot use the number of slashes “/” to determine how legitimate a site is.

3.3.2 URL Length more than 100 characters

The length of an address is another characteristic by which a phishing address can often be identified. Although legitimate URLs are short and easy to memorize but things become hard when the URL strings are more than 100 characters. It has been observed that attackers use this trick to deceive users by using very long URLs in order to hide the true address of a

phishing webpage. Overtime, legitimate sites especially email and shopping websites now use long URL's for their product.

Examples:

Phishing URL:

http://s70.n57.n84.n66.static.myhostcenter.com/QYFLWHKACZ8OY0LFL7UCGS84HKGQ17J4XOPECJSTP7NQU2DQ8XOFIFQ6AEMUYU0KZSQPVNA20AG342B5K2UCAUAVH42EUC1Z67G/6R1HALA19LIDZMU671S625H3IU64854OQJQN3S08LP8DIUSL78E9GVLPL01R4IFLMNWU8MMUMZM78QNZI47SC52DXDTL6TT1RX9/dados_ibk5.php?00009gC7fIU64-cC0S8ioewcqqd%2017ovjq6i1?op=7000493=2012333
<http://floridagunsupply.com/skin/install/pdf/home/service/costumer/information/check/88acf4c35af6f99de8983409c86374aa/index/web/d1689ee77a90617fc3a2e3b7c8abf6c3/login.php>

Legitimate URL:

<https://www.jumia.com.ng/rovite-transparent-skeleton-men-mechanical-watch-black-gold-4350356.html>

It is worthy to know that the URL above is a product page of a wrist watch on www.jumia.com. It is now difficult to also use URL string length to determine how legitimate a website is. A phishing site can also put up a fake link with a long URL to a product, and ask the user to put down important information.

The anti-phishing algorithm below is used by the proposed scheme to decide the status of a URL in an e-mail message:

```

Begin
Select Email type
Load Email
If login details (username and password) is
in Email
    Login to Email
Else
    Flag Error
If Login Successful
    Display Email
Load URL Link embedded in the email
Check URL Link for malicious activity.
If Malicious activity detected
    Alert the user and update the system
behaviour

```

Else
Continue user session
End

3.2 Alerter

This is the last component of the email based phishing detector which communicates the results of the rules analysis from the Detector engine to the users. The Alerter uses a strong warning language to make users reason with the consequences of their action if the detection result returns negative (i.e. phishing sites).

4. Implementation and Evaluation

Our proposed phishing detection app is implemented using Java. The apps automatically synchronized with the Yahoo and Gmail e-mail app on a mobile phone where it is installed. The app takes the URL on an opened e-mail message and checks for its legitimacy using the detection rules. If the suspicious URL in the email message comes out as a phishing one, the system gives warning to the user.

The proposed solution was installed and tested on 5 Android mobile devices, devices ranging from different software specifications i.e. android versions, memory sizes, and screen sizes. It was also tested using an android emulator (Bluestack and Memu) on a machine with powerful hardware to fully test the proposed solution. To evaluate the performance of the proposed app, a dataset consisting of 200 legitimate websites from Alexa and 200 Phish websites from PhishTank were used. The following definitions are necessary for understanding of our evaluation results:

True Positive (TP): The phishing pages that classified correctly as phishing.

False Negative (FN): The phishing pages that classified incorrectly as legitimate.

True Negative (TN): The legitimate pages that classified correctly as legitimate.

False Positive (FP): The legitimate pages that classified incorrectly as phishing.

Our evaluation results are presented in Figure 3. The app returned a True Positive of 94.5% and True Negative of 100% on the collected dataset.

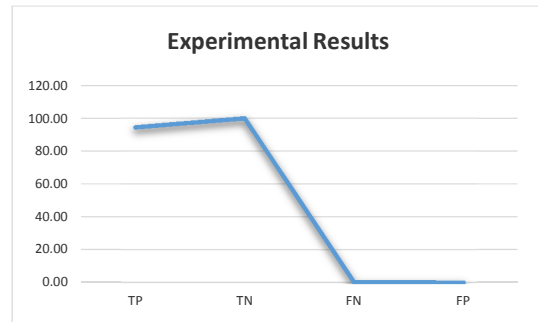


Figure 3: Experimental Results

5. Conclusion and Future Works

There has been an aggregative rise in phishing attacks in the past couple of years. This rise has remained unabated with the proliferation of mobile devices which makes many unsuspecting users vulnerable to phishing attacks. One way in which mobile devices have assisted most users is in the area of email management. However, most email message now comes with embedded suspicious URLs. In this work, an E-mail app is designed and implemented to assist mobile users to detect fraudulent URLs in their email messages. Our experimental results indicated that the proposed app is very effective in detecting phishing URLs with negligible overheads on the device memory.

Moreover, since the proposed system is rule-based, it has capability to detect newer attacks where the detection rules are applied. In the future, we hope to extend the proposed solution by adding more rules without unnecessary memory overheads. The extension should support more email providers and provide ability to check email contents for phishing detection.

This work makes the following contributions to anti-phishing research:

- It eliminates the difficulty of navigating URL from one application to another by providing email services (Gmail and Yahooemail)
- Protect users against phishing on mobile platforms using certain URL characteristic.
- Provide a secured email system.

References

- Aaron, G., & Rasmussen, R. (2015). *Global phishing survey 2H2014: Trends and domain name*.
APWG. (2016). *APWG Security Report*.

- Arun, V. (2016). Computers in Human Behavior Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks. *Computers in Human Behavior*, 198-207.
- Chiew, L., Chang, H., Sze, N., & Tiong, K. (2015). Utilization of website logo for phishing detection. *Computers and Security*.
- Hong, J. (2012). *The State of phishing attacks*. ACM.
- Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing Detection: A Literature Survey. *IEEE*.
- Orunsolu, A., Bamgboye, O., Alaran, M., Sodiya, A., & Omorinola, A. (2015). A Users' Awareness Study of Anti-phishing Security Tips. *Proceedings of the 2nd ICET*. Ilisan Remo: Department of Computer Science, Babcock University.
- RSA. (2014). *Anti-Fraud Command Center*. RSA monthly online fraud report.
- Shahriar, H., Klintic, T., & Clincy, V. (2015). Mobile Phishing Attacks and Mitigation Techniques.
- Tan, C., Chiew, K., & Sze, S. (2017). Phishing Webpage Detection Using Weighted URL Tokens for Identity Keywords Retrieval. 133-139.

About the Authors

Orunsolu A. Abiodun is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, Nigeria. He

currently undergoes his Ph.D degree at the Federal University of Agriculture, Abeokuta. His research interests include Network Security, Cryptographic protocols, Information retrieval and Intelligent systems.

Alaran M. Adunni is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, Nigeria. She is presently on her Ph.D degree at the Federal University of Agriculture, Abeokuta where her research interest is on Similarity measurement. She is a member of Nigeria Computer Society and Computer Professional Registration Council of Nigeria.

Sakiru O Kareem is a Senior Technologist in the Department of Computer Science, Moshood Abiola Polytechnic, Nigeria. He holds a Master degree in Information Communication Technology (MICT) from Federal University of Agriculture, Abeokuta. He is a member of Computer Professional Registration Council of Nigeria.

Ayobami G. Oke is a graduate of Computer Science from Moshood Abiola Polytechnic, Nigeria. He has a good background in VB.NET programming language. He specializes in developing mobile applications with emphasis on their security considerations.

Adejonwo E.I is a graduate of Computer Science from Moshood Abiola Polytechnic, Abeokuta.

Implication of Variation in Ant Colony Optimization Algorithm (ACO)
Parameters on Feature Subset Size

Ronke S. Babatunde, Rafiu M. Isiaka, Oyeranmi J. Adigun, Akinbowale N. Babatunde 1

Reengineering open source software for extended functionality:
The case of Open Medical Record System (OpenMRS)

Iwara I. Arikpo and Anthony T. Okoro 9

A Mobile-Based Phishing Detector App for E-Mails

Abdul A. Orunsolu, Misturah A. Alaran, Sakiru O. Kareem, Ayobami G. Oke, Eniola I. Adejonwo 19