African Journal of Computing & ICTs

Volume 7. No. 3. September, 2014

September, 2014
www.ajocict.net

A Journal of the Institute of Electrical & Electronics Engineers (IEEE)
Computer Chapter Nigeria Section

ISSN- 2006-1781

www.ajocict.net

CONTENTS

Editorial Board

www.ajocict.net

Foreword

The African Journal of Computing & ICT remains at the nexus of providing a platform for contributions to discourses, developments, growth and implementation of Computing and ICT initiatives by providing an avenue for scholars from the developing countries and other nations across the world to contribute to the solution paradigm through timely dissemination of research findings as well as new insights into how to identify and mitigate possible unintended consequences of ICTs. Published papers presented in this volume provide distinctive perspective on practical issues, opportunities and dimensions to the possibilities that ICTs offer the African Society and humanity at large. Of note are the increasing multi-disciplinary flavours now being demonstrated by authors collaborating to publish papers that reflect the beauty of synergistic academic and purpose-driven research. Obviously, these developments will drive growth and development in ICTs in Africa.

The Volume 7, No. 3, September 2014 Edition of the African Journal of Computing & ICTs contains journal articles with a variety of perspective on theoretical and practical research conducted by well-grounded scholars within the sphere of computer science, information systems, computer engineering, electronic and communication, information technology and allied fields across the globe. While welcoming you to peruse this volume of the African Journal of Computing and ICTs, we encourage you to submit your manuscript for consideration in future issues of the Journal

We welcome comments, rejoinders, replication studies and notes from readers.

Wishing you a productive reading

Thank you

Longe Olumide Babatope PhD
Managing Editor
Afr J Comp & ICTs
September, 2014

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Grid Scheduling Algorithm Based on Generalized Knapsack Model

D. R. Aremu & S.A. Oluwasogo
Department of Computer Science.
Faculty of Communication and Information Sciences
University of Ilorin, Ilorin, Nigeria
draremu2006@gmail.com, samueloluwasogo@yahoo.com

ABSTRACT

Grid is a collaborative problem-solving environment in which one or more user jobs can be submitted without knowing where the resources are or even who own the resources. One motivation of Grid computing is to aggregate the power of widely distributed resources, and provide non-trivial services to users. To achieve this goal, efficient Grid scheduling algorithms are fundamentally important. The aim of this paper is to present a scheduling algorithm for assigning a task to the most appropriate resource on the grid such that the resource is optimally utilized subject to the condition that the capacity of the resource is not exceeded. The methodology adopted to achieve this aim are: (i) to study the existing Grid schedulers, and scheduling approaches; (ii) to design a Grid Scheduling algorithm based on a generalize assignment problems of the Knapsack Model; (iii) to use Dynamic Programming approach as a solver for the resulting resource allocation problem. The time complexity of the resulting solution algorithm was reduced to O(n*W).

Keywords: Grid Computing, Scheduling, Quality of Service, Knapsack Problem, Resource Allocation.

African Journal of Computing & ICT Reference Format:
D. R. Aremu & S.A. Oluwasogo (2014). Grid Scheduling Algorithm Based on Generalized Knapsack Model.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp1-10.

1. INTRODUCTION

The Grid is emerging as a wide-scale, distributed computing infrastructure that promises to support resource sharing and coordinated problem solving in dynamic, multi-institutional Virtual Organization [2]. Grid Resources [3] fall into the categories of computation (i.e. a machine sharing its CPU), storage (i.e. a machine sharing its RAM or disk space), communication (i.e. sharing of bandwidth or a communication path), software and licenses and special equipment (i.e. sharing of devices). One motivation of Grid computing is to aggregate the power of widely distributed resources, and provide non-trivial services to users. To achieve this goal, efficient Grid scheduling algorithms are fundamentally important. A scheduler is needed to locate the computers on which to run an application, and to assign the jobs required. This can be as simple as taking the next available resource, but often this task involves prioritizing job queues, managing the load, finding idle machines. The Grid scheduling algorithm is the crux of the Grid scheduler. A large variety of Grid scheduling algorithms have been proposed in the literature, but no consideration has been given to Knapsack model as a possible scheduling algorithm for grid resource allocation. The aim of this paper is to present a scheduling algorithm for assigning a task to the most appropriate resource on the grid such that the grid resource is optimally utilized subject to the condition that the capacity of the resource is not exceeded.

The methodology adopted to achieve this aim are: (i) to study the existing Grid scheduling algorithm; (ii) to design a Grid Scheduling algorithm based on a generalize assignment problems of the Knapsack Model; (iii) to use Dynamic Programming approach as a solver for the resulting resource allocation problem. The rest of the paper is organized as follows: section 2 presents the related work; section 3 discusses the generalize assignment problem of the Knapsack model; section 4 presents the Dynamic Programming approach for solving the generalize assignment problem of the Knapsack model; section 5 discuss the simulation of the presented model; while section 6 concludes the paper.

2. RELATED WORK

As scheduling is usually a process to find optimal solutions, several analogies from natural phenomena have been introduced to form powerful heuristics, which have proven to be highly successful. Some of the common characteristics of Nature's heuristics are the close resemblance to a phenomenon existing in nature, namely, non-determinism, the implicit presence of a parallel structure and adaptability [1]. Abraham et al [1] and Braun et al [14] present three basic heuristics implied by Nature for Grid scheduling, namely, *Genetic Algorithm* (GA), *Simulated Annealing* (SA) and *Tabu Search* (TS), and heuristics derived by a combination of these three algorithms.

2.1 Genetic Algorithm (GA)

GA is an evolutionary technique for large space search. The general procedure of GA search is as follows [14]:

1) Population generation: A population is a set of *chromosomes* and each represents a possible solution, which is a mapping sequence between tasks and machines. The initial population can be generated by other heuristic algorithms, such as Min-min (called seeding the population with a Min-min chromosome).

2) Chromosome evaluation: Each chromosome is associated with a fitness value, which is the makespan of the task-machine mapping this chromosome represents. The goal of GA search is to find the chromosome with optimal fitness value.

3) Crossover and Mutation operation: Crossover operation selects a random pair of chromosomes and chooses a random point in the first chromosome. For the sections of both chromosomes from that point to the end of each chromosome, crossover exchanges machine assignments between corresponding tasks. Mutation randomly selects a chromosome, then randomly selects a task within the chromosome, and randomly reassigns it to a new machine.

4) Finally, the chromosomes from this modified population are evaluated again. This completes one iteration of the GA. The GA stops when a predefined number of evolutions is reached or all chromosomes converge to the same mapping. This genetic algorithm randomly selects chromosomes. Crossover is the process of swapping certain sub-sequences in the selected chromosomes. Mutation is the process of replacing certain sub-sequences with some task-mapping choices new to the current population. Both crossover and mutation are done randomly. After crossover and mutation, a new population is generated. Then this new population is evaluated, and the process starts over again until some stopping criteria are met.

The stopping criteria can be, for example,
1) no improvement in recent evaluations;
2) all chromosomes converge to the same mapping;
3) a cost bound is met. For its simplicity, GA is the most popular Nature's heuristic used in algorithms for optimization problems. In the realm of Grid scheduling, we can find other examples in, [3], [14].

2.2 Simulated Annealing (SA)

SA is a search technique based on the physical process of annealing, which is the thermal process of obtaining low-energy crystalline states of a solid. At the beginning, the *temperature* is increased to melt the solid. If the temperature is slowly decreased, particles of the melted solid arrange themselves locally, in a stable "ground" state of a solid. SA theory states that if temperature is lowered sufficiently slowly, the solid will reach thermal equilibrium, which is an optimal state.

By analogy, the thermal equilibrium is an optimal task-machine mapping (optimization goal), the temperature is the total completion time of a mapping (cost function), and the change of temperature is the process of mapping change.
If the next temperature is higher, which means a worse mapping, the next state is accepted with certain probability. This is because the acceptance of some "worse" states provides a way to escape local optimality which occurs often in local search [14].

A SA algorithm is implemented in [14]. The initial system temperature is the makespan of the initial (random) mapping. The initial SA procedure implemented here is as follows. The first mapping is generated from a uniform random distribution. The mapping is mutated in the same manner as the GA, and the new makespan is evaluated. If the new makespan is better, the new mapping replaces the old one. If the new makespan is worse (larger), a uniform random number $z \in [0, 1]$ is selected. Then, z is compared with y, where If $z>y$, the new (poorer) mapping is accepted; otherwise it is rejected, and the old mapping is kept. So, as the system temperature "cools", it is more difficult for poorer solutions to be accepted. This is reasonable because when the temperature is lower, there is less possibility to find a better solution starting from another poorer one. After each mutation, the system temperature is reduced to 90% of its current value. (This percentage is defined as the cooling rate.) This completes one iteration of SA. The heuristic stops when there is no change in the makespan for a certain number of iterations or the system temperature approaches zero. Another example of SA's application in the Grid can be found in [14].

2.3 Tabu Search (TS)

TS is a meta-strategy for guiding known heuristics to overcome local optimality and has now become an established optimization approach that is rapidly spreading to many new fields. The method can be viewed as an iterative technique which explores a set of problem solutions, denoted by X, by repeatedly making moves from one solution s to another solution s' located in the neighbourhood $N(s)$ of s. These moves are performed with the aim of efficiently reaching an optimal solution by minimizing some objective function $f(s)$. In [14], TS is implemented beginning with a random mapping as the initial solution, generated from a uniform distribution. To manipulate the current solution and move through the solution space, a *short hop* is performed. The intuitive purpose of a short hop is to find the nearest local minimum solution within the solution space. The basic procedure to perform a short hop is to consider, for each possible pair of tasks, each possible pair of machine assignments, while the other assignments are unchanged. If the new makespan is an improvement, the new solution is saved, replacing the current solution.

The short hop procedure ends when (1) every pair-wise remapping combination has been exhausted with no improvement, or (2) the limit on the total number of successful hops (*limithops*) is reached. When the short hop procedure ends, the final mapping from the local solution space search is

added to the tabu list. The tabu list is a method of keeping track of the regions of the solution space that have already been searched. Next, a new random mapping is generated, and it must differ from each mapping in the tabu list by at least half of the machine assignments (a successful *long hop*).
The intuitive purpose of a long hop is to move to a new region of the solution space that has not already been searched. After each successful long hop, the short hop procedure is repeated. The stopping criterion for the entire heuristic is when the sum of the total number of successful short hops and successful long hops equals *limithops*. Then, the best mapping from the tabu list is the final answer.

2.4 Combined Heuristics

GA can be combined with SA and TS to create combinational heuristics. For example, The Genetic Simulated Annealing (GSA) [1] heuristic is a combination of the GA and SA techniques. In general, GSA follows procedures similar to the GA outlined above. However, for the selection process, GSA uses the SA cooling schedule and system temperature and a simplified SA decision process for accepting or rejecting a new chromosome.

These Nature's heuristics were only relatively introduced into the scheduling area and more work needs to be done to fit them in a Grid context. There are a lot of interesting questions [14]. First, the meaning of controlling measurements in such heuristics needs to be refined. For example, in SA, the cooling rate is usually set to a fixed value to control the number of iterations the search process will perform; the question is: how to choose this value as well as the threshold value *y* for accepting a poorer mutation? Second, there is a trade-off between the search cost and the degree of optimality of solutions found. For example, in a genetic algorithm, historical knowledge can be used to guide the chromosome selection, crossover, or mutation process so that the search process can converge quickly. But this adjustment seems to contradict the philosophy of an evolutionary algorithm: randomization and diversity generate better results, and it may not bring a better solution. Third, what is the effect of the dynamic nature of the Grid on these algorithms? More practical experiments should be done to find proper answers to these questions.

3. MODEL FOR RESOURCE ALLOCATION ALGORITHM

The General Assignment Problem (*GAP)* is best described using knapsack problems [2]. Given *n set of* tasks to be done on the Grid and *m Grid* resources, with p_{rj} as the profit associated with assigning task *j* to resources *r*, w_{rj} as the weight of assigning task *j* to resources *r*, and c_r as the capacity of resource *r*, assign each task *j* to exactly one resource *r*, not exceeding resource capacities.

Defining x_{rj} = 1 if task j is assigned to resource r; 0 otherwise

The *GAP* can be formulated as:

$$\text{Maximize} \quad z = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{rj} x_{rj} \qquad (1)$$

$$\text{Subject to} \quad \sum_{j=1}^{n} w_{rj} x_{rj} \leq c_r \quad r \in M = \{1, \ldots, m\} \quad \{\text{respect knapsack capacities}\} \qquad (2)$$

$$\sum_{j=1}^{n} x_{rj} = 1, \quad j \in N = \{1, \ldots, n\} \quad \{\text{all tasks assigned}\} \qquad (3)$$

$xij \in \{0,1\}$ *for all* $i \in M$ *and* $j \in N$

where x_{rj} = 1 if task j is allocated to resource i; 0 otherwise

4. DYNAMIC PROGRAMMING APPROACH TO KNAPSACK MODEL

Using Dynamic Programming approach for the above knapsack problem

Step 1: Decompose the problem into smaller problems.
We construct array Z[0..m, 0..n], for $0 \leq r \leq m$, $0 \leq j \leq n$, the entry Z[r,j] will store d maximum (combined) total profit of items {1,2,3…,j} of combined size at most C_r.
If we compute all the entries of this array, then the array entry Z[M,N] will contain the maximum total profit of items that can be assigned into the knapsack/resources, that is, the sub-solution to our problem.

Step 2: Recursively define the value of an optimal solution in terms of solutions to smaller problems.

Initial Settings: Set Z[r, 0] = 0, for $0 \leq r \leq m$, no item
Illegal case

Z[r, j] = -∞, for r < 0,

Recursive Step:
Use Z[r, j] = max (Z[r, j-1], P[r, j] + Z[r-1, j-1]) for $1 \leq j \leq n$, $0 \leq r \leq m$ if and only if (W[i][j] < C[r]) and C[r] = C[r] – W[r]);

Step 3: Bottom-up computing Z[r, j] using iteration. **Bottom:**
Z[r, 0] = 0 for all $0 \leq r \leq m$
Computing of table using Z[r, j] = max (Z[r, j-1], P[r, j] + Z[r-1, j-1]) row by row

www.ajocict.net

Table 1 Tabular Representation of Dynamic Programming Concept

Z[r,j]	J = 0	1	2	3	...	...	N
r = 0	0	0	0	0	...	...	0
1	→	→	→	→	→	→	→
2	→	→	→	→	→	→	→
: .	→	→	→	→	→	→	→
M	→	→	→	→	→	→	→

Up ↓ Bottom

5. IMPLEMENTATION

The implementation of the generalized Knapsack Scheduling algorithm using the Dynamic programming approach was carried out with theoretical and computation experiments for two scenarios. The result for each of the scenario is presented in sub sections 5.1, 5.2, 5.3, and 5.4 below:

5.1 Theoretical Experiment For Scenario 1

The algorithm presented for solving Dynamic Programming approach to Knapsack problem as stated in section 4 above was experimented with Scenario 1 given below.

Scenario 1:
Given, the value of n (set of tasks) is 4; m (set of Grid resources) is 5.

C_r (Capacity of the Grid resources) =

R	1	2	3	4	5
C_r	50	40	45	30	60

Table 2: Weight of task j if assigned to resource r

W_{rj}	j = 1	2	3	n = 4
r = 1	5	6	7	8
2	3	12	4	5
3	6	7	9	10
4	4	8	6	2
m = 5	2	11	14	13

Table 3: Profit of task j if assigned to resource r.

P_{rj}	j = 1	2	3	n = 4
r = 1	20	15	60	70
2	50	40	30	20
3	10	15	20	25
4	15	40	55	50
m = 5	60	65	10	5

To compute the optimal solution by applying the Dynamic programming approach to the problem stated in the scenario 1 above, we construct an array Z[0..5, 0..4], for $0 \leq r \leq m$ and $0 \leq j \leq n$ such that the maximum of the profits of tasks {1, 2, 3, 4}and capacity C_r

Our interest is to find Z[r, j] = max (Z[r, j-1], P[r, j] + Z[r-1, j-1]) - for $1 \leq j \leq n$, $0 \leq r \leq m$

if and only if (W[i][j] <= C[r]) and (p[r][j] + Z[r-1][j-1] > Z[r][j-1]);

That is we must be sure that the W[i][j] is not greater than capacity C[r]

Case when: r= 0, j = 1 , Z[0,j] = 0 for all j member of N
r = 1, j = 1, Z[1,1] = max { 0, 20 + 0} > 0? = 20
r = 1, j = 2, Z[1,2] = max { 0, 15 + 0} > 20? = 20
r = 1, j = 3, Z[1,3] = max {0, 60 + 0} > 20? = 60
r = 1, j = 4, Z[1,4] = max {0, 70 + 0} > 60? = 70
r = 2, j = 1, Z[2,1] = max {0, 50 + 0} > 0? = 50
r = 2, j = 2, Z[2,2] = max {0, 40 + 20} > 50? = 60
r = 2, j = 3, Z[2,3] = max {0, 30 + 20} > 60? = 60
r = 2, j = 3, Z[2,4] = max {0, 20 + 60} > 50? = 80

If we continue in that manner, we obtain a table as shown below for Z[r,j]
.

Table 4: Result of computed maximum profit for using resource r for task j

Z[r,j]	**j = 0**	**1**	**2**	**3**	**4**
r = 0	**0**	**0**	**0**	**0**	**0**
1	**0**	**20**	**20**	**60**	**70**
2	**0**	**50**	**60**	**60**	**80**
3	**0**	**10**	**65**	**80**	**85**
4	**0**	**45**	**50**	**120**	**130**
5	**0**	**60**	**110**	**110**	**125**

Also we obtain the corresponding values for X_{rj} as shown in the table below;

The value X_{rj} = 1 if the task j is assign to the resource r.

Table 5: Result of computed status of resourcr r if used for task j

X_{rj}	j = 0	1	2	3	4
r = 0	0	0	0	0	0
1	0	1	0	1	1
2	0	1	1	0	1
3	0	1	1	1	1
4	0	1	1	1	1
5	0	1	1	0	1

The optimal solution for Z[r,j] is obviously 130. from table 4.

5.2 Computational Experiment For Scenario 1

The result of the computational experiment for scenario 1 using C++ programming language for computation is as shown below:

Read in the N set of tasks to be done: 4

Read in the m Grid of knapsacks or resources to store items: 5

Read in the capacity of knapsack: 1 50
Read in the capacity of knapsack: 2 40
Read in the capacity of knapsack: 3 45
Read in the capacity of knapsack: 4 30
Read in the capacity of knapsack: 5 65

Read in the weight of assigning task 1to knapsack: 1 5
Read in the weight of assigning task 2to knapsack: 1 6
Read in the weight of assigning task 3to knapsack: 1 7
Read in the weight of assigning task 4to knapsack: 1 8

Read in the weight of assigning task 1to knapsack: 2 3
Read in the weight of assigning task 2to knapsack: 2 12
Read in the weight of assigning task 3to knapsack: 2 4
Read in the weight of assigning task 4to knapsack: 2 5

Read in the weight of assigning task 1to knapsack: 3 6
Read in the weight of assigning task 2to knapsack: 3 7
Read in the weight of assigning task 3to knapsack: 3 9
Read in the weight of assigning task 4to knapsack: 3 10

Read in the weight of assigning task 1to knapsack: 4 4
Read in the weight of assigning task 2to knapsack: 4 8
Read in the weight of assigning task 3to knapsack: 4 6
Read in the weight of assigning task 4to knapsack: 4 2
Read in the weight of assigning task 1to knapsack: 5 2
Read in the weight of assigning task 2to knapsack: 5 11
Read in the weight of assigning task 3to knapsack: 5 14
Read in the weight of assigning task 4to knapsack: 5 13

Read in the cost(profit) of assigning task 1to knapsack: 1
20
Read in the cost(profit) of assigning task 2to knapsack: 1
15
Read in the cost(profit) of assigning task 3to knapsack: 1
60
Read in the cost(profit) of assigning task 4to knapsack: 1
70

Read in the cost(profit) of assigning task 1to knapsack: 2
50
Read in the cost(profit) of assigning task 2to knapsack: 2
40
Read in the cost(profit) of assigning task 3to knapsack: 2
30
Read in the cost(profit) of assigning task 4to knapsack: 2
20

Read in the cost(profit) of assigning task 1to knapsack: 3
10
Read in the cost(profit) of assigning task 2to knapsack: 3
15
Read in the cost(profit) of assigning task 3to knapsack: 3
20
Read in the cost(profit) of assigning task 4to knapsack: 3
25

Read in the cost(profit) of assigning task 1to knapsack: 4
15
Read in the cost(profit) of assigning task 2to knapsack: 4
40
Read in the cost(profit) of assigning task 3to knapsack: 4
55
Read in the cost(profit) of assigning task 4to knapsack: 4
50

Read in the cost(profit) of assigning task 1to knapsack: 5
60
Read in the cost(profit) of assigning task 2to knapsack: 5
65
Read in the cost(profit) of assigning task 3to knapsack: 5
10
Read in the cost(profit) of assigning task 4to knapsack: 5 5

The capacity of knapsack: 60 50 55 40 45

The weight of assigning task to knapsack:

5 6 7 8
4 12 4 5
6 7 9 10

4 8 6 2
2 11 14 13

The cost of assigning task to knapsack:

20 15 60 70
50 40 30 20
10 15 20 25
15 40 55 50
60 65 10 5

The result table for the maximum profit computed is:

0 0 0 0 0
0 20 20 60 70
0 50 60 60 80
0 10 65 80 85
0 45 50 120 130
0 60 110 110 125

The result table for the X(r,j) is:

0 0 0 0 0
0 1 0 1 1
0 1 1 0 1
0 1 1 1 1
0 1 1 1 1
0 1 1 0 1

The maximum total computed Profit of 6 items that can be assigned to the 5 knapsack is: 130

Optimal Solution is at Position: Z[5,5]

5.3 Theoretical Experiment for Scenario 2

The algorithm presented for solving Dynamic Programming approach to Knapsack problem as stated in section 4 above was experimented with Scenario 2 given below:

Scenario 2: Given, the value of n (set of tasks) is 6, m (set of Grid resources) is 5.

C_r (Capacity of the Grid resources) =

r	1	2	3	4	5
Cr	60	50	55	40	45

Array W_{rj} showing the weight of task j if assigned to resource r

w_{rj}	j = 1	2	3	4	5	6
r=1	4	5	8	12	15	14
2	14	6	7	5	9	13
3	7	8	19	10	16	5
4	6	8	10	12	7	20
5	7	9	11	13	4	3

Array P_{rj} showing the profit of task j if assigned to resource r

P_{rj}	j = 1	2	3	4	5	6
r=1	30	25	15	20	15	10
2	40	50	17	25	35	40
3	55	75	30	45	40	35
4	25	30	35	40	45	50
5	50	55	60	65	70	80

Array Z[r,j] showing the maximum profit obtain if resource r is assign to task j

Z[r,j]	j = 0	1	2	3	4	5	6
r=0	0	0	0	0	0	0	0
1	0	30	30	30	30	30	30
2	0	40	80	80	80	80	80
3	0	55	115	115	125	125	125
4	0	25	85	150	155	155	155
5	0	50	80	145	215	255	225

The value X_{rj} = 1 if the task j is assign to the resource r.

X_{rj}	j = 0	1	2	3	4	5	6
r=0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0
2	0	1	1	0	0	0	0
3	0	1	1	0	1	0	0
4	0	1	1	1	1	0	0
5	0	1	1	1	1	1	0

5.4 Computational Experiment For Scenario 2

Output for the Computation Experiment 2

Read in the N set of tasks to be done: 6

Read in the m Grid of knapsacks or resources to store items: 5

Read in the capacity of knapsack: 1 60
Read in the capacity of knapsack: 2 50
Read in the capacity of knapsack: 3 55
Read in the capacity of knapsack: 4 40
Read in the capacity of knapsack: 5 45

Read in the weight of assigning task 1to knapsack: 1 4
Read in the weight of assigning task 2to knapsack: 1 5
Read in the weight of assigning task 3to knapsack: 1 8
Read in the weight of assigning task 4to knapsack: 1 12
Read in the weight of assigning task 5to knapsack: 1 15
Read in the weight of assigning task 6to knapsack: 1 14

Read in the weight of assigning task 1to knapsack: 2 14
Read in the weight of assigning task 2to knapsack: 2 6
Read in the weight of assigning task 3to knapsack: 2 7
Read in the weight of assigning task 4to knapsack: 2 5
Read in the weight of assigning task 5to knapsack: 2 9
Read in the weight of assigning task 6to knapsack: 2 13

Read in the weight of assigning task 1to knapsack: 3 7
Read in the weight of assigning task 2to knapsack: 3 8
Read in the weight of assigning task 3to knapsack: 3 19
Read in the weight of assigning task 4to knapsack: 3 10
Read in the weight of assigning task 5to knapsack: 3 16
Read in the weight of assigning task 6to knapsack: 3 5

Read in the weight of assigning task 1to knapsack: 4 6
Read in the weight of assigning task 2to knapsack: 4 8
Read in the weight of assigning task 3to knapsack: 4 10
Read in the weight of assigning task 4to knapsack: 4 12
Read in the weight of assigning task 5to knapsack: 4 7
Read in the weight of assigning task 6to knapsack: 4 20

Read in the weight of assigning task 1to knapsack: 5 7
Read in the weight of assigning task 2to knapsack: 5 9
Read in the weight of assigning task 3to knapsack: 5 11
Read in the weight of assigning task 4to knapsack: 5 13
Read in the weight of assigning task 5to knapsack: 5 4
Read in the weight of assigning task 6to knapsack: 5 3

Read in the cost(profit) of assigning task 1to knapsack: 1 30
Read in the cost(profit) of assigning task 2to knapsack: 1 25
Read in the cost(profit) of assigning task 3to knapsack: 1 15
Read in the cost(profit) of assigning task 4to knapsack: 1 20
Read in the cost(profit) of assigning task 5to knapsack: 1 15
Read in the cost(profit) of assigning task 6to knapsack: 1 10

Read in the cost(profit) of assigning task 1to knapsack: 2 40
Read in the cost(profit) of assigning task 2to knapsack: 2 50
Read in the cost(profit) of assigning task 3to knapsack: 2 17
Read in the cost(profit) of assigning task 4to knapsack: 2 25
Read in the cost(profit) of assigning task 5to knapsack: 2 35
Read in the cost(profit) of assigning task 6to knapsack: 2 40

Read in the cost(profit) of assigning task 1to knapsack: 3 55
Read in the cost(profit) of assigning task 2to knapsack: 3 75
Read in the cost(profit) of assigning task 3to knapsack: 3 30
Read in the cost(profit) of assigning task 4to knapsack: 3 45
Read in the cost(profit) of assigning task 5to knapsack: 3 40
Read in the cost(profit) of assigning task 6to knapsack: 3 35

Read in the cost(profit) of assigning task 1to knapsack: 4 25
Read in the cost(profit) of assigning task 2to knapsack: 4 30
Read in the cost(profit) of assigning task 3to knapsack: 4 35
Read in the cost(profit) of assigning task 4to knapsack: 4 40
Read in the cost(profit) of assigning task 5to knapsack: 4 45
Read in the cost(profit) of assigning task 6to knapsack: 4 50

Read in the cost(profit) of assigning task 1to knapsack: 5 50
Read in the cost(profit) of assigning task 2to knapsack: 5 55
Read in the cost(profit) of assigning task 3to knapsack: 5 60
Read in the cost(profit) of assigning task 4to knapsack: 5 65
Read in the cost(profit) of assigning task 5to knapsack: 5 70
Read in the cost(profit) of assigning task 6to knapsack: 5 80

The capacity of knapsack: 60 50 55 40 45

The weight of assigning task to knapsack:

4	5	8	12	15	14
14	6	7	5	9	13
7	8	19	10	16	5
6	8	10	12	7	20
7	9	11	13	4	3

The cost of assigning task to knapsack:

30	25	15	20	15	10
40	50	17	25	35	40
55	75	30	45	40	35
25	30	35	40	45	50
50	55	60	65	70	80

The result table obtain for the maximum profit computed is:

0	0	0	0	0	0	0
0	30	30	30	30	30	30
0	40	80	80	80	80	80
0	55	115	115	125	125	125
0	25	85	150	155	155	155
0	50	80	145	215	225	225

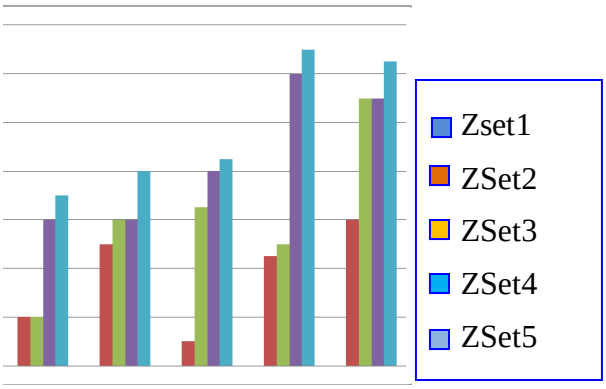
The result table obtained for the X(r,j):

0	0	0	0	0	0	0
0	1	0	0	0	0	0
0	1	1	0	0	0	0
0	1	1	0	1	0	0
0	1	1	1	1	0	0
0	1	1	1	1	1	0

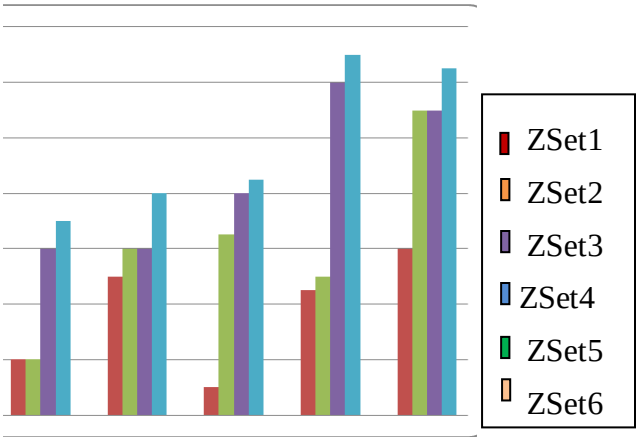
The maximum total computed Profit of 6 items that can be assigned to the 5 knapsack is: 225. Optimal Solution is at Position: Z[5,5]

6. DISCUSSION OF RESULT

The graphical interpretation of the results from both the computation and theoretical experiments for the two scenarios as previously discussed are shown in graphs 1 to 4 below.

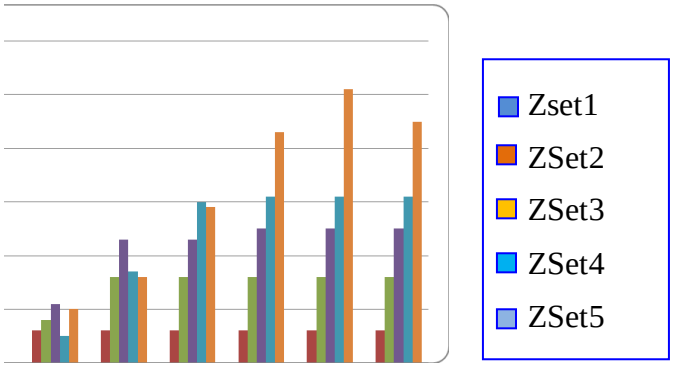


Graph 1: Representing value Z[r,j] for theoretical experiment1 with Profit (Z(r,j) plotted against resources (r).

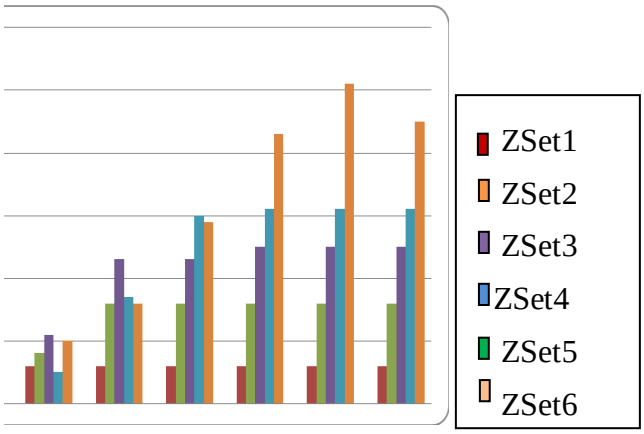


Graph 2: Representing value Z[r,j] for Computational experiment1 with Profit (Z(r,j) plotted against resources (r).

From Graph 1 and Graph 2 showing the result of the theoretical and computational experiments respectively, the optimal value is 130. It can be deduced that the knapsack model is appropriate and accurate for allocating resources on the Grid.



Graph 3: Representing value Z[r,j] for theoretical experiment 2 with Profit (Z(r,j) plotted against resources(r).



Graph 4: Representing value Z[r,j] for Computational experiment 2 with Profit (Z(r,j) plotted against resources(r).

The maximum total computed Profit of 6 items that can be assigned to the 5 knapsack is: 225 Optimal Solution is at Position: Z[5,5].

From Graph 1 and Graph 2 showing the comparative result of the theoretical and computational experiments respectively, it can be deduced that the knapsack model is appropriate for allocating Grid resources. The Optimal value from the two graphs is 130.

7. SUMMARY AND CONCLUSION

This study aim at presenting an efficient scheduling algorithm for allocating grid resources in an optimize manner. In order to achieve this aim, the study adopted a Generalized Knapsack Problem scheduling algorithms which uses Dynamic problem approach to compute the mapping of job to an appropriate resource. The study compared results of the theoretical and computational experiments of the resource allocation models.

8. FINDINGS

- Dynamic programming is a useful technique for solving certain kind of problems
- This study showed that the Generalised assignment problem of the Knapsack Model can be used for resource allocation on the grid.
- The running time of the knapsack programming is better than other algorithm discussed in the related work.
- Running time of dynamic programming algorithm vs. naive algorithm for Generalized Knapsack problem is O(W*n) vs. O(2^n) [20]

- Generalised Knapsack Problem scheduling algorithms uses Dynamic problem approach to compute the mapping of job to an appropriate resource. This increases the performance of a scheduler and in turn the Grid environment.

In conclusion, the presented Generalize Assignment Problem of Knapsack Problem can used to solve the problem of selecting an appropriate grid resource for a set of task on the grid without underutilizing the resource and also not exceeding the resource capacity. This increases the performance of a scheduler and in turn the Grid environment.

REFERENCES

[1] Kleinrock, L. *MEMO on* Grid Computing, University of California, Los Angeles, 1969. [2] Foster, I., Kesselman, C. and Tuecke, S. *The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of High Performance Computing applications*, Vol. 15, No.3, pp. 200-222, 2001.

[3] Satyarth Kumar Singh, Efficient Grid Scheduling Algorithm based on Priority Queues Computer science and engineering department Thapar university Patiala – 147004

[4] Yanmin ZHU, *A Survey of Grid Scheduling,* Department of Computer Science Hong Kong University of Science and Technology.

[5] Foster, I. and Kesselman, C. *The Grid: Blueprint for a New Computing Infrastructure.*Morgan Kaufmann, San Francisco, CA, 677 pp, 1999.

[6] Ferreira, L., Bieberstein, N., Berstis, V., Armstrong, J., *Introduction to Grid Computing with Globus*, Redbook IBM Corp.

[7] Comparison of Grid Computing vs. Cluster Computing at "http://www.jatit.org/research/introduction_grid_computing.htm".

[8] "CORBA: Common Object Request Broker Architecture, http://www.omg.org/gettingstarted/corbafaq.htm.

[9] Gregor von Laszewski, Manish Parashar, Snigdha Verma, Jarek Gawor, Kate Keahey, Nell Rehn**,** "A CORBA Commodity Grid Kit", The Applied SoftwareSystemsLaboratory, Department of Electrical and Computer Engineering.

[10] Domenico Talia, Paolo Trunfio, "A P2P Grid Services-Based Protocol: Design and Evaluation", University of Calabria.

[11] *Jean-Christophe Durand, Grid Computing: A Conceptual and Practical Study*, Universite de Lausanne, Nov 8, 2004.

[12] Yan Liu, "Grid Scheduling", *Department of Computer Science, University of Iowa.*49

[13] *N. Tonellotto, Information Science and Technologies Institute Italian National Research Council Italy, R. Yahyapour Institute for Robotics Research University of Dortmund Germany, "*A Proposal for a Generic Grid Scheduling Architecture*".*

[14] Fangpeng Dong and Selim G. Akl, *Scheduling Algorithms for Grid Computing: State of art and Open Problems*, Technical-Report No. 2006-504, School of Computing,Queen's University, Kingston, Ontario, Canada, January 2006.

[15] <http://www.cs.wisc.edu/condor/>

[16] Abraham Silberschatz, Peter Baer Galvin, Greg Gagne, "Operating System Principles Seventh Edition".

[17] <www.csse.monash.edu.au/~davida/ /nimrodg.htm>

[18] T. Casavant, and J. Kuhl, *A Taxonomy of Scheduling in General-purpose Distributed Computing Systems*, in IEEE Trans. on Software Engineering Vol. 14, No.2, pp. 141--154, February 1988.

[19] http://www.cse.unl.edu/~goddard/ /CSCE310J

[20] Silvano Martello and PaoloToth. Knapsack Problems: Algorithms and Computer Implementation. John Wiley & Sons Ltd., 1090. ISBN 0-471-92420-2.

[21] Hewlett-Packard Laboratories. Generalized Knapsack solvers for Multi-Unit Combinatorial Auctions: Analysis and application to Computational Resource Allocation. February 15th, 2004.

Authors' Short Biography

Dr. Aremu Dayo Reuben is a lecturer at the Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin, Ilorin, Nigeria. He obtained a BSC (Hons.) degree in Mathematics at the University of Ilorin, Ilorin, Nigeria in 1989, a Master Degree in Mathematics with option in Computer Science at the University of Ilorin, Ilorin in 1998 and a PhD Degree in Computer Science from the University of Zululand, South Africa in 2008. His research focused on Software Engineering, Grid Computing and Information security, He can be reached by phone on +2348033608919 and through E-mail draremu2006@gmail.com

Mr. Samuel Oluwasogo is a computer programmer/Analyst at the Information and communication Technology Centre, the Federal Polytechnic, Offa, Nigeria. He obtained a BSC Computer Science at the University of Ilorin, Ilorin, Nigeria in 2004 and a Master of Science Degree in Computer Science at the University of Ilorin, Ilorin, Nigeria in 2013. His research has focused on Grid Scheduling Algorithm Models. He can be reached by phone on +2348032488366 and through E-mail samueloluwasogo@gmail.com

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Forged Signature Detection Using Artificial Neural Network

T. O. Oladele, K. S. Adewole & A. O. Oyelami
Department of Computer Science
University of Ilorin
Ilorin, Nigeria.
tinuoladele@gmail.com, adewole.ks@gmail.com, 08walexy@gmail.com
+234 8164916082 +234 8038544348 +234 8038341776

T. N. Abiodun
Department of General Studies
Nigerian Institute of Jouralism
Ogba Estate, Ikeja, Lagos.
abitess@live.com
+234 8054571566

ABSTRACT

Crimes and corruptions are practices that gradually cripple the economy of a nation most especially in Nigeria. Nigerian government has strived hard to reduce these acts perpetrated by the citizens. This is evident in the struggles of Economic and Financial Crime Commission (EFCC) and Independence Corrupt Practices and other Related Offences Commission (ICPC) to reduce frauds in both public and private sectors due to signature forgery which attempts to commit financial crimes and other related offences. Forged signature is an illegal copy of signature that looks like a genuine signature usually used for financial fraud. Identity verification (authentication) in computer systems has been traditionally based on something that one has such as key, magnetic or chip card or that one knows such as PIN or password. Things like keys or cards, however, tend to get stolen or lost and passwords are often forgotten or disclosed. In this paper, a neural network algorithm was employed to develop a system that can verify and detect forged signatures. The effect of the signature verification and detection is discussed and its impact on the economy is highlighted. Result of the proposed Java application shows its capability in detecting forged signatures. The system has the capability to learn from previous data and to assist EFCC and ICPC in detecting and investigating fraudulent activities.

Keywords: Neural Network, Algorithm, Biometrics, Signature & Forgery..

African Journal of Computing & ICT Reference Format:
T. O. Oladele, K. S. Adewole, T. N. Abiodun & A. O. Oyelami (2014). Forged Signature Detection Using Artificial Neural Network. Afr J. of Comp & ICTs. Vol 7, No. 3. Pp11-20.

1. INTRODUCTION

The purpose of this paper is to provide a fast, safe and easy to use application to detect forged signatures and this can be achieved by using several techniques but in this paper, neural network was employed being one of the techniques of Data mining. Data mining is the process of analyzing and extracting knowledge from large volume of data. Data mining can also be called Knowledge Discovery from Data (KDD). There is always need for signature verification to be carried out before any payment is made, say in banks, but this is done manually by making comparison between the signature presented and the one in the database in which only eyes cannot detect the authenticity of the signature. The signature of a person is an important biometric attribute of a human being which can be used to authenticate human identity [1, 2]. As signatures continue to play a very important role in financial, commercial and legal transactions, truly secured authentication becomes more and more crucial.

Handwritten signatures are considered as the most natural method of authenticating a person's identity. A signature by an authorized person is considered to be the "seal of approval" and remains the most preferred means of authentication. However, human signatures can be handled as an image and recognized using computer vision and neural network as a technique of data mining. With modern computers, there is need to develop fast algorithms for signature recognition [2]. This application will analyze, compare, mine, and evaluate patterns then present knowledge of the mined data. In a broad sense data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories.

Signature is a special case of handwriting which includes special characters and flourishes. Many signatures can be unreadable. They are a kind of artistic handwriting objects. However, a signature can be handled as an image, and hence, it can be recognized using computer vision and artificial neural network techniques. Signature recognition and verification involves two separate but strongly related tasks: one of them is identification of the signature owner, and the other is the decision about whether the signature is genuine or forged [3]. Considering the importance of Data mining, it is pertinent to study the manual method of signature verification. This paper dealt with the problem of verifying the authenticity of signatures manually and avoid the problems which occur when carried out manually. Some of the drawbacks of the existing system were identified and a computerized system that will be compatible with the existing system and be more user-friendly was designed. In this paper, the back propagation training algorithm of neural network was employed in the development of a system that can verify and detect forged signatures. The Forged Signature Detection System was implemented by using Java programming language.

2. RELATED WORKS

A number of biometric methods have been introduced, but few have gained wide acceptance. This approach, concentrates on some techniques to gather knowledge. Adrian Perrig introduces the BiBa signature scheme. BiBa stands for Bins and Balls signature. A collision of balls under a hash function in bins forms the signature [4]. BiBa signature scheme is a new signature construction that uses one-way functions without trap doors. The most important features of BiBa signature scheme is a low verification overhead and a relatively small signature size. In comparison to other one-way function based signature schemes, BiBa has smaller signatures and is at least twice as fast to verify (which probably makes it one of the fastest signature scheme to date for verification). Ihler et al presented modeling dynamical processes. They used this methodology to model handwriting stroke data, specifically signatures, as a dynamical system and show that it is possible to learn a model capturing their dynamics for use either in synthesizing realistic signatures and in discriminating between signatures and forgeries even though no forgeries have been used in constructing the model [5]. Another study reported in [6] explained the method for on-line signature authentication, which is based on an event-string modeling of features derived from pen-position and pressure signals of digitizer tablets. The intra- class variability is one of the key problems in biometrics.

Pivonka, and Nepevny, proposed a new controller by combination of Generalized Predictive Control (GPC) algorithm and Neural Network model, with many advantages. Neural model has the ability to observe system changes and adapted itself, therefore regulator based on this model also will be adaptive. Algorithm was implemented in MATLAB-Simulink with aspect of future implementation to Programmable Logic Controller (PLC).

It was tested on mathematical and physical models in soft-real-time realization. Predictive controller in comparison with classical PSD controller and the advantages and disadvantages were shown [7]. Plamondon and Lorette made an assertion that most of the work in off-line forgery detection has been on random or simple forgeries and less on skilled or simulated forgeries [8]. Papadimitrioy and Terzidis, proposed a fuzzy system that approximated the accurate set of rules keeping only the more important aspects of the data. The approximation algorithms either received an a priori description of a set of fuzzy sets or, especially for the case when interpretable fuzzy sets could not be pre-specified by the experts, the algorithm presented for building them automatically. After the construction of the interpretable fuzzy partitions, the developed algorithms extract from the SVFI rules a small and concise set of interpretable rules.

Finally, the Pseudo Outer Product (POP) fuzzy rule selection ordered the interpretable rules by using a Hebbian like evaluation in order to present the designer with the most capable rules [9]. Unluturk et al, conducted an extensive study on the use of Neural Network in Biometric researches. They developed the emotion recognition neural network (ERNN) to classify the voice signals for emotion recognition [10]. Dule, et al, worked on outdoor vehicle images to determine the color of the vehicle and color classes. In this work, the Performances of different feature sets obtained by various color spaces and different classification methods were taken into account in order to improve the outdoor vehicle color recognition [11].

2.1 Types Of Forgeries

The main task of any signature verification system is to detect whether the signature is genuine or counterfeit. Forgery is a crime that aims at deceiving people. Since actual forgeries are difficult to obtain, the instrument and the results of the verification depend on the type of the forgery [1, 12]. Basically there are three types that have been defined:

1. **Random forgery**: this can normally be represented by a signature sample that belongs to a different writer i.e. the forger has no information whatsoever about the signature style and the name of the person.
2. **Simple forgery**: this is a signature with the same shape or the genuine writer's name.
3. **Skilled forgery**: this is signed by a person who has had access to a genuine signature for practice [1, 8, 13].

2.2 Overview of Neural Networks

Neural networks offer a mathematical model that attempts to mimic the human brain. Knowledge is represented as a layered set of interconnected processors, which are called neurons. Each node has a weighted connection to other nodes in adjacent layers. Individual nodes take the input received from connected nodes and use the weights together with a simple function to compute output values. Learning in neural networks is accomplished by network connection weight changes while a set of input instances is repeatedly passed through the network.

Once trained, an unknown instance passing through the network is classified according to the values seen at the output layer. Surveys existing work on neural network construction, attempting to identify the important issues involved, directions the work has taken and the current state of the art. Neurons only fire when input is bigger than some threshold. It should, however, be noted that firing doesn't get bigger as the stimulus increases, it is an all or nothing arrangement. A typical neuron collects signals from others through a host of fine structures called dendrites.

The neuron sends out spikes of electrical activity through a long, thin stand known as an axon, which splits into thousands of branches. At the end of each branch, a structure called a synapse converts the activity from the axon into electrical effects that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning occurs by changing the effectiveness of the synapses so that the influence of one neuron on another changes [1, 14]. Figure 1 shows a typical diagram of a neuron.

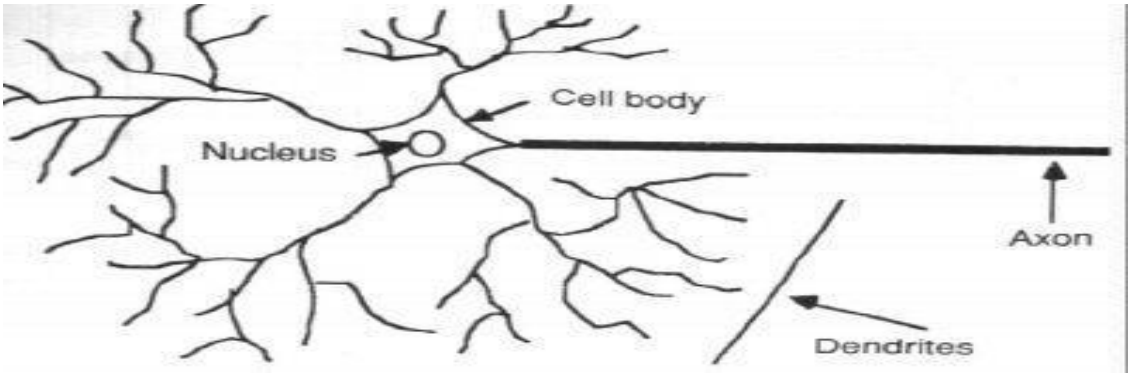


Figure 1: Diagram of a Biological Neuron.

The human brain, for example, contains approximately 1011 neurons, each connected on average to 10, 000 other neurons, making a total of 1,000,000,000,000,000 = 1015 synaptic connections. Neural networks represent an attempt at a very basic level to imitate the type of nonlinear learning that occurs in the networks of neurons found in nature. McCulloch and Pitts [15] are generally recognized as the designer of the first neural network.

They combined many simple processing units together that could lead to an overall increase in computational power. They suggested many ideas like: a neuron has a threshold level and once that level is reached the neuron fires. It is still the fundamental way in which artificial neural network operates. The McCulloch and Pitts's network had a fixed set of weights. The artificial neuron model was introduced by McCulloch-Pitts and it is known as Threshold Logic Unit. The diagram of an artificial neuron is shown in figure 2.

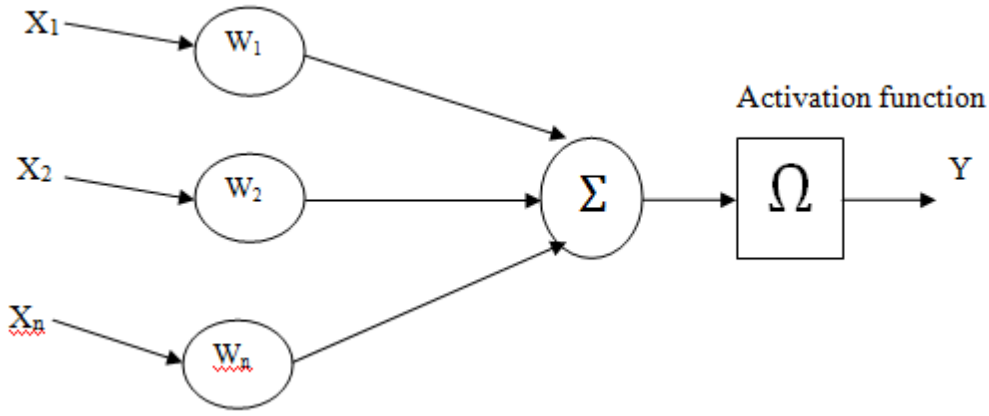


Figure 2: Diagram of an Artificial neuron.

3. METHODOLOGY

This involves the specification of procedures for collecting and analyzing data necessary to define or solve the problem. The scope of this paper covers the technique used to detect forged signature and type of classification approach that can be used to solve the problem of detecting forged signature is neural network. According to figure 2, a set of input connections brings in activations from other neurons. Then a processing unit sums the inputs, and then applies a non-linear activation function and the output line transmits the result to other neurons. The values $w_1, w_2, \ldots w_n$ are weights to determine the strength of input vector $x = [x_1, x_2, \ldots x^T]$. Each input is multiplied by the associated weight of the neuron connection $x^T w$. the positive weight excites and the negative weight inhibits the node output. The equation can be given thus:

$$I = x^T.w = x_1w_1 + x_2w_2 + \ldots x_nw_n = \sum_{i=0}^{n} xiwi \quad (1)$$

The node's internal threshold Ω is the magnitude offset.
It affects the activation of the node output Y as:

$$Y = f(I) = f\{ \sum_{i=0}^{n} xiwi - \Omega_k \} \quad (2)$$

To generate the final output Y, the sum is passed on to a non-linear filter f called the activation function which relates the output Y . Neuron generally do not produce an output unless their total output goes above a threshold value. The total input for each neuron is the sum of the weighted inputs to the neuron minus its threshold value. The is then passed through the sigmoid function. The equation for a transition in neuron is :

$$a = 1/(1 + \exp(-x)) \quad (3)$$

Where

$$x = \sum_i a_i wi - Q$$

a is the activation for the neuron
a_i is the activation for neuron $_i$
w_i is the weight
Q is the threshold subtracted.

The activation function f performs a mathematical operation on the signal output. The most common activation functions are; linear function, piecewise linear function, tangent hyperbolic function, threshold function, sigmoid (S shaped) function. Computer understands and process data in binary that is zeros and ones. The methodology used discusses the pre-processing performed, the signature database, and the neural network features. Pre-processing in neural network offers the following pre-processing method and they are background elimination, width normalization, and thinning.

Signature database: since the application is a standalone application, it does not require database but it uses the traditional form of database system which is the file system containing all the scanned signatures and they are encrypted in such a way that only the application can access it for security measures. The application is designed to recognize signature in which the signatures are taken in four places for system to train them and recognize it after which it is train by the because that is the main essence why neural network is used because it can learn and uses what it learnt to do comparison with the new signature that would presented and verification is done by comparing. Training is the act of presenting the networks with some sample data and modifying the weight to better approximate the desired function and that is how the weight comes in.

The weight in a neural network is the most important factor in determining its function. There are two main types of training and they are [1]:

1. Supervised Training: In this type of training, it supplies the neural network with inputs and the desired outputs and response of the network to the input is measured. Also the weights are modified to reduce the difference between the actual and desired outputs.
2. Unsupervised Training: It only supplies the inputs then the neural network adjusts its own weight so that similar inputs cause similar outputs. The network identifies the patterns and differences in the inputs without any external assistance.

3.1 The Preprocessing Step

The system will take in signature by either scanning it in to the system or written on a stylus, after which pre-processing is done and under pre-processing, three steps will take place in this stage and they are:

a) Background Elimination: Many image processing applications require differentiation of objects from the image background. Thresholding is the most trivial and easily applicable method for this purpose. It is widely used in image segmentation. Threshold technique is used for differentiating the signature pixels from the background pixels. In this type of application, interest is on dark objects on a light background and therefore a threshold value *T* called the brightness threshold is appropriately chosen and applied to image pixels. After the thresholding, the pixels of the signature would be 1 and the other pixels which belong to the background would be 0.

b) Width Normalization: Irregularities in the image scanning and capturing process may cause signature dimensions to vary. Furthermore, height and width of signatures vary from person to person and sometimes even the same person may use different size signatures. First there is the need to eliminate the size differences and obtain a standard signature size for all signatures. During the normalization process, the aspect ratio between the width and height of a signature is kept intact and after the

process, all the signatures will have the same dimension. For example width normalization is calculated by:

$$Normalized\ area = \frac{\text{Signature Area}}{\text{Area enclosed in bounding box}}$$

c) Thinning: The goal of thinning is to eliminate the thickness differences of pen by making the image one pixel thick. The comparison is done based on what the system learnt from training. The signature presented is flagged genuine with percentage if and only if the signature is within the range of the already trained signature existing in the system. But if the signature presented is not within the range of the existing signature, it is flagged forge with the percentage of its forgery. The implementation of the system can be achieved with Java program alongside with Netbeans which will serve as editor, compiler and it will also allow the program modules to run at the same time and will make it to be a standalone application. Basically, the design of the system is divided into two stages:

A. Training stage: A training stage consist of four major steps:
1. Retrieval of a signature image from a database (file system)
2. Image pre-processing
3. Feature extraction
4. Neural network training

B. Testing stage: A testing stage consists of five major steps :
1. Retrieval of a signature to be tested from a database
2. Image pre-processing
3. Feature extraction
4. Application of extracted features to a trained neural network
5. Outputting in percentage.

4.1 User Interface

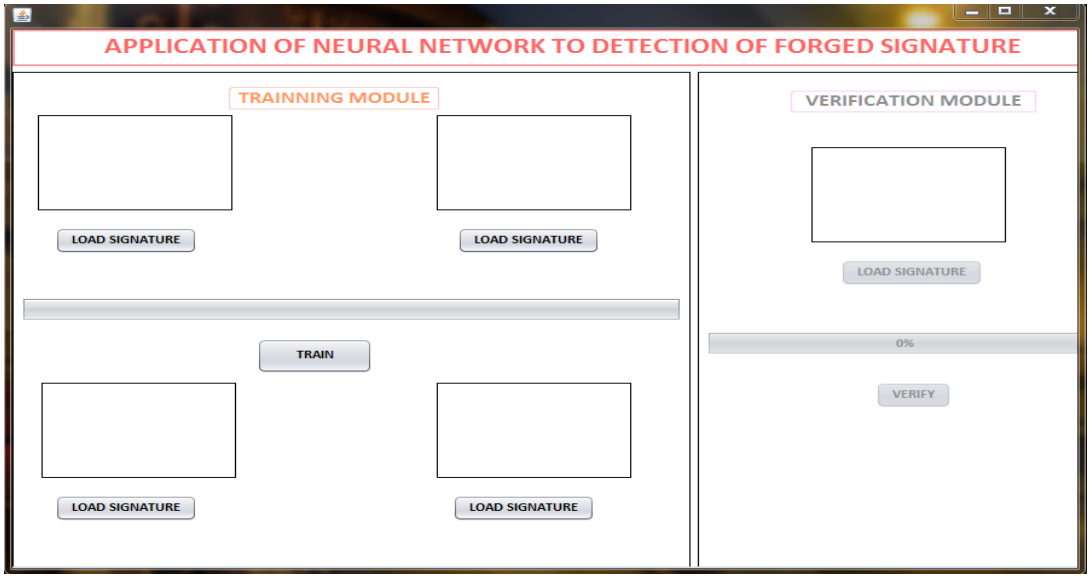


Figure 3: User Interface

4. RESULTS

Below are a few screen shots or stages of the application consisting of two separate verified signatures with different percentage of authenticity. Interfaces or stages highlighted include the User interface, the Training stage, the verification stage and the Testing stage with fake signature. All these results make up the workings of the Signature Verification System.

The User Interface is shown in figure 3. This is the first page the user of the application sees. It features an interactive map of the signature detection application.

4.2 Training Stage of the First Sample Signatures

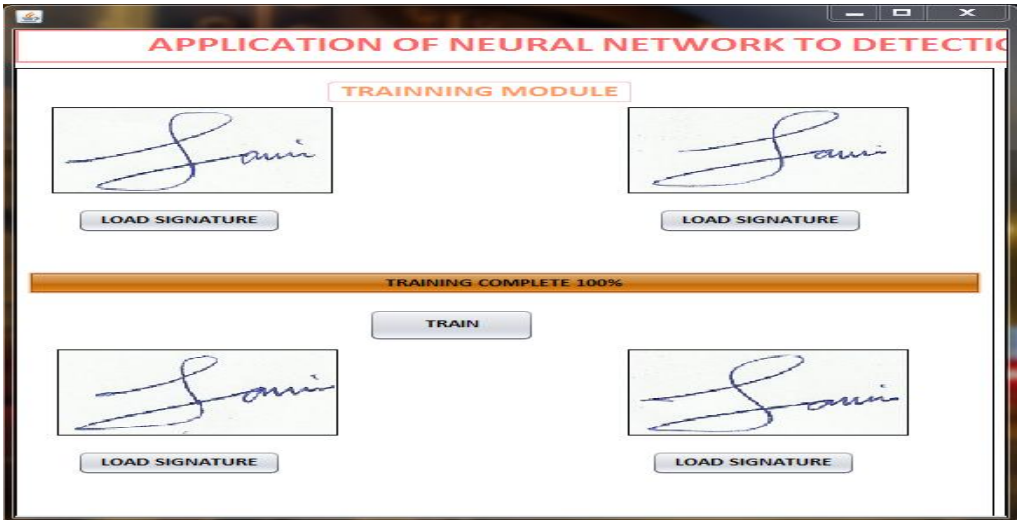


Figure 4: Training Stage

The Training stage is the module where signatures are uploaded into four places and trained so that the application can recognize it, convert the signatures to binary and store them into a file as shown in figure 4.

4.3 Verification Stage of the First Sample Signatures

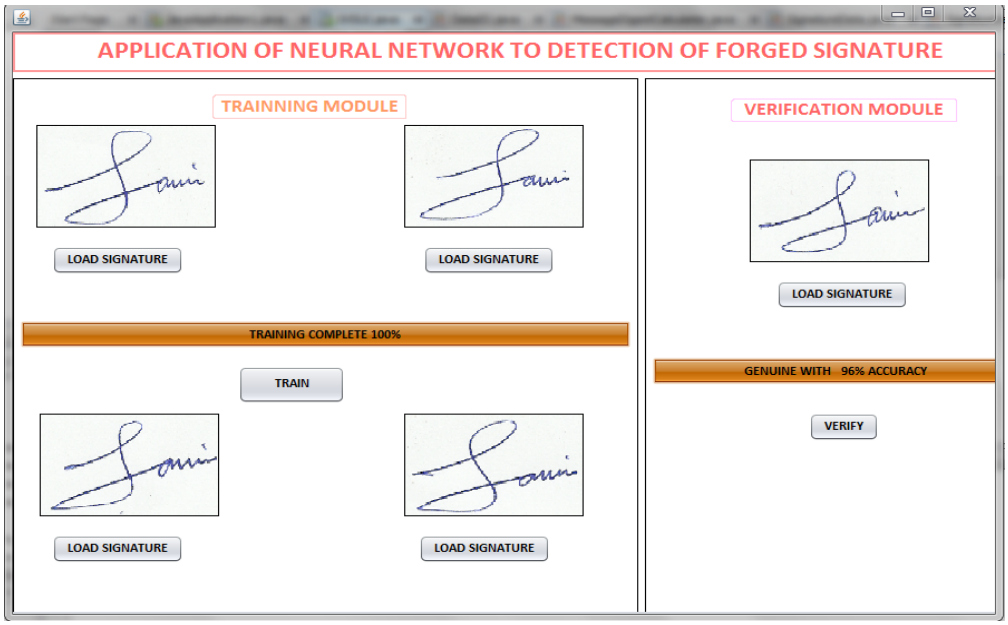


Figure 5: Verification Stage

This stage verifies the signature trained by the application and comparison is made between the signature presented and the already existing ones in the database. This now tells the accuracy of the signature. The signature in figure 5 shows 96% accuracy which is considered genuine for authenticating the user.

4.4 Training Stage of the Second Sample Signatures

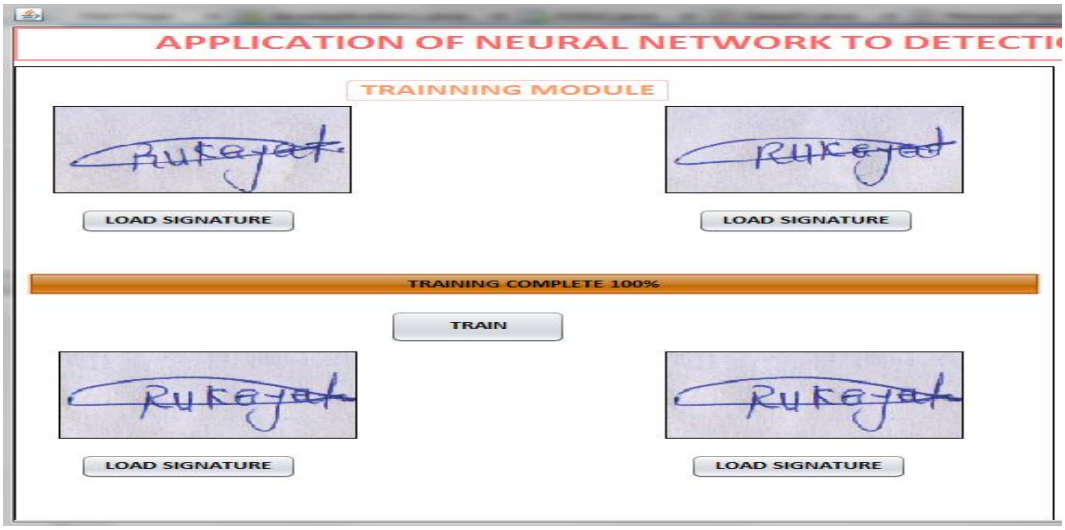


Figure 6: Training Stage

This is also the module where signatures are uploaded in four places and trained so that the application can recognize it, convert the signatures to binary and store them into a file as shown in figure 6.

4.5 Verification Stage of the Second Sample Signatures

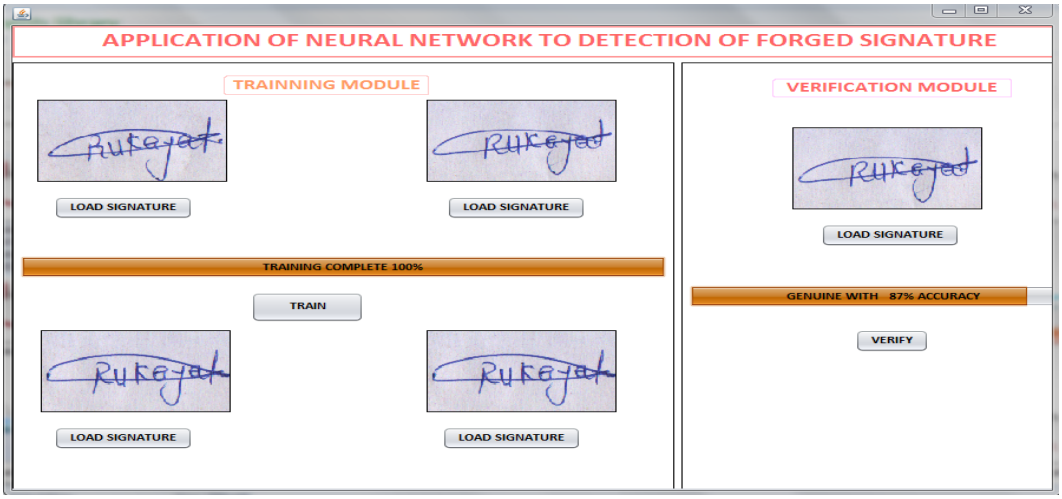


Figure 7: Verification Stage

This stage verifies the signature trained by the application and comparison is made between the signature presented and the already existing ones. This now tells the accuracy of the signature. The verification module indicates 87% similarity between the signed signature and the one in the database as shown in figure 7. This can also be considered for authenticating the user.

5. DISCUSSION OF RESULTS

Figures 4 and 6 shows the training stage for the application to learn the pattern of the signature and save them into database for comparison during verification. It can be seen that the training percentage will reach hundred and these denote completion of training after which verification can later be done by the application. Figure 5 denotes the verification of the first signature, it is seen that the percentage of the signature does not reach 100%, the reason being that the owner of the signature can never sign the exact way he or she has signed before and this has led to the variation in the percentage of the accuracy of the signature. The signature in the verification module can only give 100% if and only if the signature to be verified was taken from the signature used for training the application. In the case of figure 7 which is the verification stage of the second signature, the accuracy of the signature also did not reach 100% because of little variation when the owner is signing such as the speed, pen used, state the owner is when signing and so on. It can be concluded that, the accuracy of the signature can never be 100%.

6. CONCLUSION

In this paper, a signature verification system was been developed by applying neural network. The efficacy of the system was tested on a large database of signatures. The designed system can be used as an effective signature verification system. In achieving this, neural network is used and it is implemented with Java. This paper presents a method for Forged Signature Detection. The extracted features are used to train a neural network by using the back propagation training algorithm. The network could classify all genuine and forged signatures correctly. Technologies used in implementing the signature verification application include the Unified Modeling Language (UML), Java Virtual Machine (JVR), NetBeans compiler and JQuery.

REFERENCES

[1] Abikoye, O. C., Mabayoje, M. A., & Ajibade, R. (2011). Offline Signature Recognition and Verification using Neural Network. International Journal of Computer Applications. Vol. 35, No. 2, 0975-8887.

[2] Bhatia, M. (2013). Off-line Hand Written Signature Verification using Neural Network. International Journal of Application Innovation in Engineering Management (IJAIEM). Vol. 2 Issue 5. ISSN 2319-4847.

[3] OZ, C., Ercal, F. & Demir, Z. (2003). Signature Recognition and Verification with ANN. Third International Conference on Electrical Electronics Engineering, Bursa. Turkey.

[4] Perrig, A. (2001). The Biba One-Time Signature and Broadcast Authentication Protocol. Berkeley, University of California.

[5] Ihler, A. T., Fisher, J. W. & Willsky, A. S. (2001). Nonparametric Estimators for Online Signature Authentication. International Conference on Acoustics, Speech, and Signal Processing, USA. Vol. 6, pp. 3473-3476.

[6] Schimke, S., Vielhauer, C., & Dittmann, J. (2004). Using Adapted Levenshtein Distance For On- Line Signature Authentication. 17th International Conference on Pattern Recognition (ICPR'04), Vol. 2, pp.931-934.

[7] Pivonka, P. and Nepevny, P. (2005). Generalized Predictive Control with Adaptive Model Based on Neural Networks. Proceedings of the 6th WSEAS International Conference on NEURAL NETWORKS, Lisbon, Portugal,16-18, pp. 1-4.

[8] Plamondon, R., & Lorette, G. (1989). Automatic Signature Verification and Writer Identification: the State of the Art 1989-1993. Pattern Recognition. Vol. 22, No. 2, pp. 107-131.

[9] Papadimitrioy, S., & Terzidis, K. (2007). Classification Process Analysis of Bioinformatics Data with a Support Vector Fuzzy Inference System", Proceedings of the 8th WSEAS International Conference on Neural Networks, Vancouver, British Columbia, Canada, pp. 90-95.

[10] Unluturk, M. S., Oguz, K. & Atay, C. (2009). Emotion Recognition Using Neural Networks. Proceedings of the 10th WSEAS International Conference on NEURAL NETWORKS, Prague, Czech Republic, pp. 82-85.

[11] Dule, E., Gokmen, M., & Beratoglu, S. (2010). A Convenient Feature Vector Construction for Vehicle Color Recognition, Proceedings of the 11th WSEAS International Conference on NEURAL NETWORKS (NN '10), G. Enescu University, Iasi, Romania, pp. 250-255.

[12] Kalenova, D. (2005). Personal Authentication using Signature Recognition. www2.it.lut.fi/kurssit/03-04/010970000/seminars/Kalenova.pdf

[13] Aykanat C. et. al ,(Eds). 2004. Proceedings of the 19th International Symposium on Computer and Information Sciences, ISCIS 2004. Springer-Verlag Berlin Heidelberg New York. pp. 373-380.

[14] Stergiou, C., & Siganos, D. (2003). Neural Networks Retrieved April 1, 2011 from www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs#/report.html.

[15] McCulloch, W. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 7:115 - 133.

[16] Baradi, V. A. & Kakere, H. B. (2010). Offline Signature Recognition System. International Journal of Computer Science, 48-56.

[17] Bence, K., Benedek, T. & Hassan, C. (2009). Classification Approaches in Offline handwritten Signature Verification. WDET Transaction on Mathematics, VIII(9), 500-509.

[18] Blumenstein, M., Armand, S. & Muthukkumarasamy, (2006). Off-line Signature Verification using the Enhanced Modified Direction Feature and Neural Based Classification. International Joint Conference on Neural Networks.

[19] Dolfing, J. G., Aarts, E. H. & Van Oosterhout, J. J. (1998). On-line Signature Verification with Hidden Markov Models. Proceedings of the 14th International Conference on Pattern Recognition, Brisbane. 1309-1312.

[20] Enturk, T. S., Zgunduz, E. O., & Karshgil, E. (2005). Handwitten Signature Verification Using Image Invariant and Dynamic Features. Proceedings of the 13th European Signal Processing Conference (EUISIPCO), Antalya.

[21] Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall Publishing.

[22] Gallinari, P., Thiria, F., Badran, F., & Fogelman-Soulie, F. (1991). On the Relation between Discriminant Analysis and Multilayer Perceptron, Neural Networks. IV, 349-360.

[23] Geok, S. N., & Ong, H. S. (1993), A Neural Network Approach for Offline Signature Verification. IEEE TENCON.

[24] Gupta, G. K. (2006). The State of the art in On-line Handwritten Signature Verification. Monash University, Faculty of Information Technology, Australia.

[25] Jain, A., Griess, F. S. C. (2010). On-line Signature Verification Pattern Recognition, WSEAS Transaction on Mathematics.

[26] Manoj, K. (2012). Signature Verification Using neural Network. International Journal on Computer Science and Engineering (IJCSE). 1498-1504.

[27] McCabe, A., & Trevathan, J. (2007). Implementation and Analysis of a Handwritten Signature Verification Technique. Pp 48-58.

[28] McCabe, A., Trevathan, J. & Wayne, R. (2008). Neural Network-based Handwritten Signature Verification. Journal of Computer, School of Mathematics, Physics and Information Technology, III(8).

[29] Parker, J. R. (2005). Simple Distance between Handwritten Signatures. University of Calgary, Department of Computer Science.

[30] Plamondon, R. (1995). The Handwritten signature as a Biometric Identifier: Psychophysical Model and System Design. IEEE Conference Publication, pp. 23-27.

[31] Prashanth, C. R., Raja, K. B., Venogopal, K. R., & Patnaik, L. M. (2009). Standard Scores Correlation Offline Signature Verification System. International Conference on Advances in Computing, Control and Telecommunication Technologies.

[32] Quen-Zong, W., I-Chang, J. & Suh-Yin, L. (1997). On-line Signature Verification Using LPC Cepsrum and Neural Networks. IEEE Transactions on System, Man, and Cybernetics, pp. 148-153.

[33] Rajasekaran, S., & Vijayalakshmi, G. A. (2006). Neural Networks, Fuzzy Logic, and Genetic Algorithms", Prentice-Hall India Pvt., Sixth printing, May 2006. R. J. Vidmar (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. IEEE Trans. Plasma Sci. [Online]. 21(3). pp. 876—880.

[34] Sandeep, P., & Shailendra, D. (2011). Neural Network-based Offline Handwritten Signature Verification System using Hu's Moment Invariant Analysis. International Journal of Engineering and Advanced Technology (IJEAT) I(I), 2249-8958.

[35] Srihari, S. N., Xu, A. & Kalera, M. K. (2004). Learning Strategies and Classification Methods for Off-line Signature Verification. Proceedings of the 9th International Workshop on Frontiers in Handwritten Recognition (IWFHR-9).

[36] Sureerathanan, S., Huynh, N. P., Sureerathanan, N. & Nlkos, M. E. (2006). The Optimal Multi-layer Structure of Back Propagation Networks. Proceedings of the 7th WSEAS International Conference on Neural Networks, pp. 108-113. Cavat.

[37] Xiao, X., & Leedham, G. (1999). Signature Verification by Neural Networks with Selective Attention. Journal of Applied Intelligence, XI(2), 213-223.

[38] Zois, E., & Anastassopoulos, V. (1996). Methods for Writer Identification. Proceedings of the Third IEEE International Conference on Electronics, circuits and Systems (ICECS).

[39] Bhattacharyya, D., & Kim, T. (2010). Design of Artificial Neural Network for Handwritten Signature Recognition. International Journal of Computers and Communications. Vol. 4, Issue 3.

[40] Plamondon, R., & Leclere, F. (1994). Automatic Signature Verification: the State of the Art 1989-1993. International Journal of Recognition and Artificial Intelligence. Vol. 8, No. 3, pp. 643-660.

[41] Verma, R. & Rao, D. S. (2013). Offline Signature Verification and Identification using Angle Feature and Pixel Density Feature and both Method Together. International Journal of Soft Computing and Engineering (IJSCE). ISSN: 2231-2307, Vol. 2, Issue 4.

[42] Bhattacharyya, D., & Kim, T. ().Signature Recognition using Artificial Neural Network. Advances in Computational Intelligence, Man-Machine and Cybernetics. Pp. 183-187.

[43] Yusof, M. H. M., & Madasu V. K. (2003), Signature Verification and Forgery Detection System. Student Conference on Research Development (SCOReD) Proceeding, Purrajaya, Malaysia.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

A Multi-Gene Genetic Programming Application for Predicting Students Failure at School

J.O. Orove
Department of Computer Science
University of Port Harcourt
Rivers State, Nigeria
E-mail: joshuaorove2012@gmail.com
Tel: +234 7067992097

N.E. Osegi
Department of Information and Communication Technology
National Open University of Nigeria
Lagos State, Nigeria
E-mail: geeqwam@gmail.com
Tel: +234 7030081615
Website: www.osegi.com

B.O. Eke
Department of Computer Science
University of Port Harcourt
Rivers State, Nigeria
E-mail:bathoyol@gmail.com
Tel: +234 8037049586

ABSTRACT

Several efforts to predict student failure rate (SFR) at school accurately still remains a core problem area faced by many in the educational sector. The procedure for forecasting SFR are rigid and most often times require data scaling or conversion into binary form such as is the case of the logistic model which may lead to lose of information and effect size attenuation. Also, the high number of factors, incomplete and unbalanced dataset, and black boxing issues as in Artificial Neural Networks and Fuzzy logic systems exposes the need for more efficient tools. Currently the application of Genetic Programming (GP) holds great promises and has produced tremendous positive results in different sectors. In this regard, this study developed GPSFARPS, a software application to provide a robust solution to the prediction of SFR using an evolutionary algorithm known as multi-gene genetic programming. The approach is validated by feeding a testing data set to the evolved GP models. Result obtained from GPSFARPS simulations show its unique ability to evolve a suitable failure rate expression with a fast convergence at 30 generations from a maximum specified generation of 500. The multi-gene system was also able to minimize the evolved model expression and accurately predict student failure rate using a subset of the original expression.

Keywords: Genetic Programming, Student Failure Rate, Multi-Gene GP.

African Journal of Computing & ICT Reference Format:
J.O. Orove, N.E. Osegi & B.O. Eke (2014). A Multi-Gene Genetic Programming Application for Predicting Students Failure at School. Afr J. of Comp & ICTs. Vol 7, No. 3. Pp21-34.

1. INTRODUCTION

SFR has always being and will continue to be a major concern to stakeholders in the educational sector. It refers to the proportion (or more correctly percentage) of students not graduating in a chosen course of study [1]. It is an important aspect of educational curricula assessment as this will help educational administrators to evaluate the performance of their existing curricula, teaching system, infrastructure and student relations programmes. Since the performance of any school system is largely affected by the failure rate of the students, it becomes necessary to study this obviously very important parameter. In particular, there has been a global call to reduce the failure rates of science school students, especially in the Mathematics courses [2]. Social graphs and data mining techniques [3, 4] have been suggested and some cases. Logistic and multiple linear regression techniques have also been used to study student failures rates [5, 14]. Methodologies for investigating student failure rates or decline in academic performance using artificial intelligence techniques such as Neuro-Genetic Algorithms (NGAs), Artificial Neural Networks (ANNs), Genetic Algorithms and decision trees [6,7,8, 9], have been suggested and developed in the literature.

More recently, GP has been applied in a grammar guided genetic programming algorithm to predict if the student will fail or pass a determined course and identifies activities to stimulate its learning in a positive or negative way through the use of True Positive (TP) and True Negative (TN) concepts [10].

1.1 Statement of Problem

In our educational institutions today there is currently the need to improve student's performance by reducing the number of students who fail at school. This requirements demand that school administrators and managers employ specific tools, methods and approaches to effectively study and predict student's performance overtime. Currently, such predictive tools are either not flexible, or are very expensive and scarce. This paper seeks to fill this gap by introducing a Multi-gene genetic programming application, GPSFARPS, to facilitate the process of student failure rate analysis and to aid educational administrators predict student's failure at school more cost-effectively.

1.2 Theoretical Framework Using the GP Model

The prediction of student failure rate can be carried out using different approaches with consequently different kinds of results but with the sole purpose of evaluating the number of failing students in an exam or course of study. Models such as the Logistic Regression Model [5] have been used to predict the actual probability that a student will pass or fail a Chemistry test. However, when dealing with input data of the continuous form and with a small data set, this approach becomes ineffective leading to effect size attenuation [17]. Also the issue of coding and proper labelling has to be taken into account if the logistic technique is to be utilized as a viable predictor model.

Classification and Regression trees [18] have been used to predict students passing or failing a subject taking into consideration six socio-demographic variables (age, gender, ethnicity, education, work status, and disability). However, as the author [18] pointed out in his concluding remarks, limitations do exist with the use of classification trees which could distort results obtained in the classification process and as in the case of the logistic regression, the prevalence of small data sets leads to low prediction accuracies. More recently, educational data mining techniques [4, 6] have been applied to the growing problem of predicting student failure rate with high degree of success rates.

1.3 Research purpose

We intend to bridge the gap in this area, which is to develop a software application that can evolve a model expression for predicting student failure at school given a history of class/exam scores- called the historical data set. These model expressions can then be easily utilized by the school administrators in carrying out their predictions or failure rate calculations for a future data set.

2. RELATED LITERATURE

Prediction is one of the most frequently studied problems on Data Mining and Machine Learning researchers. It consists on predicting the value of a categorical attribute based on the values of other attributes, the predicting attributes. One of the most useful Data Mining tasks on educational data mining is classification [16]. It makes possible the prediction/classification of a student's performance, for example. There are different educational objectives for using classification, such as: discovery of potential student groups with similar characteristics and reactions to a particular pedagogical strategy[19], to detect student's "misuse or game-playing" the system which is correlated with substantially lower learning [27], to group students that are hint-driven or failure driven and find common misconceptions that students possess [20], to identify learners with a low rate of motivation and find remedial actions to lower drop-out rates [21], to classify/predict students when Intelligent Tutoring Systems (ITS) are being used [22].

To predict student outcome, some studies have been made: prediction of student's grades, on a scale from A to F, from test scores using neural networks (counter propagation networks and back propagation networks) [23]; prediction of the relevance of classes, i.e., if classes are relevant or not to student's academic success, using discriminant function analysis [24] Predicting a student's academic success (low, medium or high risk) using decision trees, random forests and neural networks [25].

Using GeSES [26], a method that has been designed specifically to work with students logs, based on C4.5 rules, a main goal was established by its authors: to test if symptoms of a bad adaptation in an adaptive course where detected. A bad adaptation was detected if the course generated a low performance, caused by bad student's inadequate environment. Genetic programming is also a broadly used method on Educational Data Mining. G3PMI [10], a grammar guided genetic programming algorithm, has been applied to predict if the student will fail or pass a determined course and identifies activities to stimulate its learning in a positive or negative way through the use of True Positive (TP) and True Negative (TN) concepts. Results show that G3P-MI achieves better performance with more accurate models than a better trade-off between such contradictory metrics as sensitivity and specificity. Also, it adds comprehensibility to the knowledge discovered and finds interesting relationships between certain tasks and the time expended to solving exercises with the final marks obtained in the course.

Tanner [28] used a k-nearest neighbour (kNN) method to predict student performance in an online course environment. Extensive experimental results from a 12-lesson course on touch-typing, with a database of close to 15000 students were presented. The results indicate that kNN can predict student performance accurately, and already after the very first lessons. They also concluded that early tests on skills can be strong predictors for final scores also in other skill-based courses. Kalles [9] used Genetic Algorithms and Decision trees for a posteriori analysis of tutoring practices based on Student's failure models. Their results showed that genetic induction of decision trees could indeed produce very short and accurate trees that could be used for explaining failures. In [10], data mining techniques and real data of about 670 middle-school students from Zacatecas, México were used to predict school failure. Several experiments were carried out in an attempt to improve accuracy in the prediction of final student performance and, specifically, students that were likely to fail.

In the first experiment the best 15 attributes was selected. Then two different approaches (Data Balancing and Cost-Sensitive approaches) were applied in order to resolve the problem of classifying unbalanced data by rebalancing data and using cost sensitive classification. The outcomes of each one of the approaches using the 10 classification algorithms and 10 fold cross validation were compared in order to select the best approach to the problem. From the results, it was deduced that OneR fared better with a TN (True Negative – Fail) rate of about 88.3% when using data balancing approach, whereas Jrip fared better with a TN rate of 93.3% with the cost sensitive approach. They examined the differential effects of prior academic achievement, psychosocial, behavioural, demographic, and school context factors on early high school grade point average (GPA) using a prospective study of 4,660 middle-school students from 24 schools and a combined Multiple Linear Regression (MLR)/ Hierarchical Linear Modeling (HLM) approach. Their findings suggest that (a) Prior grades and standardized achievement are the strongest predictors of high school GPA and (b) Psychosocial and behavioural factors (e.g., motivation, self-regulation, and social control) add incremental validity to the prediction of GPA. When comparing the relative importance of each set of predictors (the dominance analysis technique), the variance accounted for by psychosocial and behavioural factors is comparable to that accounted for by prior grades.

These findings highlight the importance of effective risk assessment based on multiple measures (i.e., academic, psychosocial, and behavioural) for the purpose of identifying risk, referring students to intervention, and improving academic success [10]. They also used a new form of genetic programming called Grammar based Genetic Programming (G3P), combined with an Interpretable Rule Classification Mining Scheme (ICRM) for the prediction of student failure rate at school using real data from high school students in Mexico.

These data were based on class marks and of high dimensions (variable intensive) and imbalanced, thus, the need for variable (dimension) reduction, cost sensitive classification/re-sampling of original datasets respectively to reduce these irregularities was emphasized. They compared three versions of their G3P/ICRM model with 10 classification algorithms to evaluate the Fail (or True Negative) rates and they found out that the best results were those obtained by their G3P/ICRM models.

However, one thing that is missing in most of the techniques is having a structured user friendly application that will facilitate the generation of a symbolic model expression for describing student failure rate over a particular period of time. Such expressions can make the job of the educational administrator easier particularly in the area of re-computation using a supportive regression approach.

3. SOFTWARE METHODOLOGY

In this section, a multi-gene genetic programming approach is employed. This approach uses an iterative radical unified modelling process wherein In particular we have used an open source framework for genetic programming GPTIPS® [11], as a basis for the developed GP application. The system is robust, modular and customizable with user friendly interfaces.

3.1 Theoretical Foundations

GP operates based on the Darwinian principle of evolution, natural selection, and survival of the fittest. Typically, GP solutions are evolved individual programs encoded in a structure referred to as a gene or gene tree. At the beginning of a GP algorithm the genes or expression trees are randomly initialized within a feasible solution space, and then they undergo reproduction, crossover and mutation. Reproduction is a technique used to replicate new genes from the original parent genes akin to sexual reproduction in the programming tournament. The programming tournament can be likened to be an environment of competing programs where the best of them is selected as the eventual winners. The two core genetic operations during reproduction are [13]:

(i) Crossover: This involves the interchange of genetic material among the solutions or genes
(ii) Mutation: This involves random changes (additions and deletions) within a gene itself.

The crossover operation employs a swapping function in software while the mutation operation uses a knock-out technique.

3.2 Genetic Programming Process

The fundamental mathematical formulations guiding a genetic programming application and corresponding pseudo-codes are not new and are given in [12]. In general a GP system seeks to minimize the mean square error of the fitted data set by evolving multiple solutions at different intervals of time. A typical architecture of a GP Programming sequence is shown in Fig1. A genetic programming process starts from providing genetic programming with basic building blocks of the solution and some method of analyzing how well a proposed solution solves the problem. This is also followed by supplying the Genetic programming with the fitness metrics which the Genetic programming (GP) will use in generating solution. In Fig 1 the process analysis of the genetic programming is presented and it shows clearly that the first set of solutions generated does not become the final solution. Instead the solutions keep evolving over and over certain number of times till an optimal solution is arrived at.

In the figure it is clear that the GP begins with some initial guess at a solution and successively attempts to improve the solution over time. Once some criterion for termination (either an ideal individual or some predefined run time) GP returns the best individual so far. That individual is deemed the result of GP or the Optimal Solution. In Fig 2 the more detail diagram shows what happens internally inside the Genetic Programming box. It begins by generating some initial population. The fitness of all individuals in the population is then assessed. It is unlikely that this initial generation will contain an ideal individual, but some will likely be better than others. We begin the GP loop by selecting the individuals that solve the problem best and allowing them to reproduce, making small random changes to their construction. As this process repeats numerous times, we find that on average, the fitness of the population tends to increase.

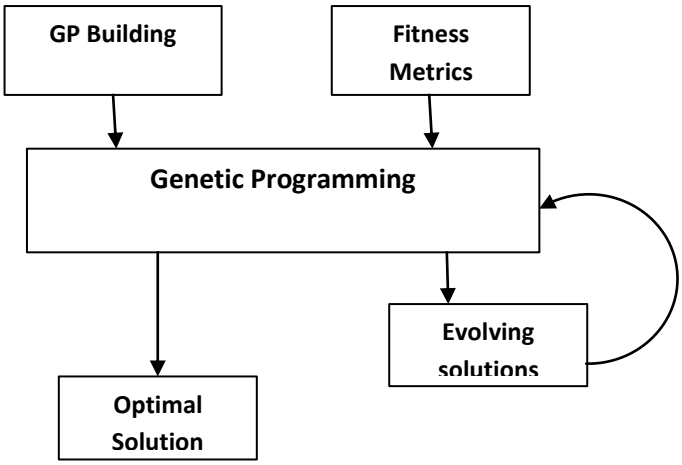


Fig 1: Simplified Genetic Programming Process Analysis

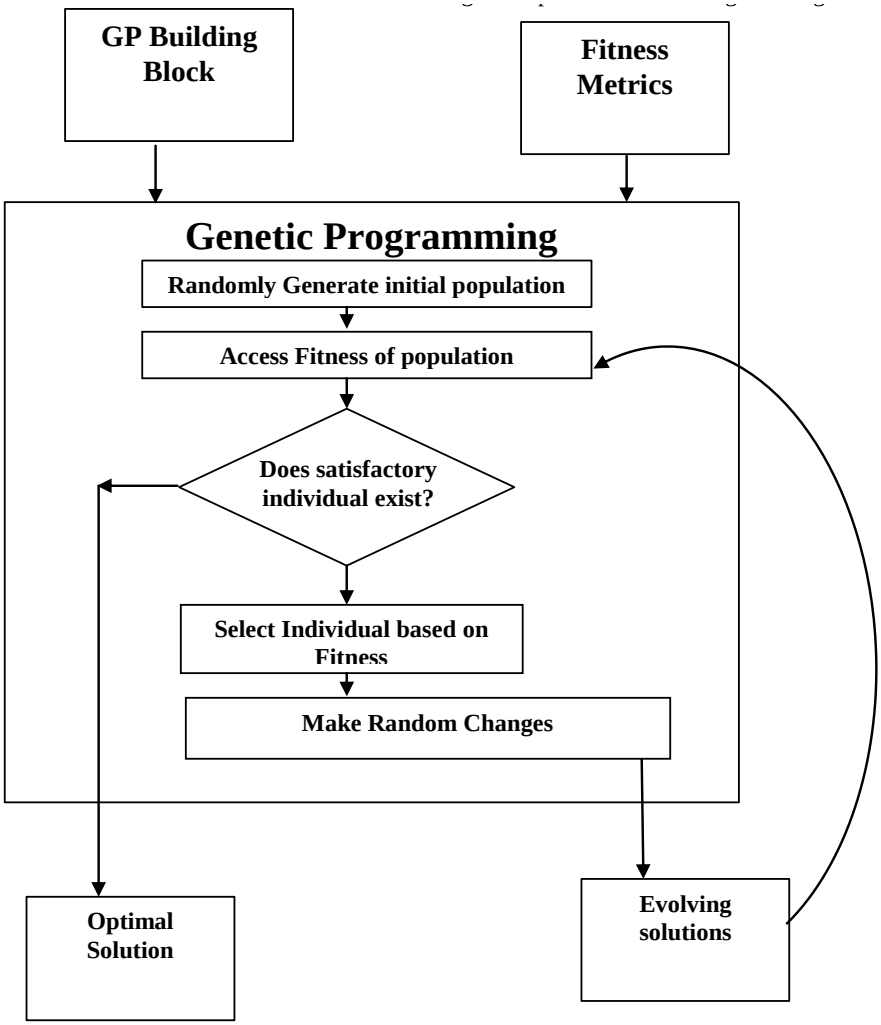


Fig 2: Main Genetic Programming Process Analysis

3.3 GP Model Solutions and Methodology

The first step in the design of a Genetic Programming is the design of the problem representation. In this section we will present the representation design of the system.

3.3.1 Representation (Building Block)

In our analysis above we have been talking a lot about individuals, for instance "Select Individual based on Fitness" but not really discussing what these individuals are made of. We know that GP aims to evolve computer programs, but what kind of computer programs? The computer programs GP evolves are programs written in some functional programming language. In functional programs, the idea is to take some input and simply return a value without dealing with computer state. You can think of functional programming more like mathematical expressions than instructions for somebody to follow. There are many functional programming languages, including LISP and OCaml. Another option is to represent the individuals using objects of whatever programming language that are been used to code the GP system. This is the way our Darwin GP Environment handles representation. However whether we choose to represent individuals in some functional programming language or with objects in memory, GP evolves individuals that can be represented as a tree structure. It is more functionally useful to consider them in this way when thinking about GP. The tree structure design in Fig 3 is the building block for our genetic programming.

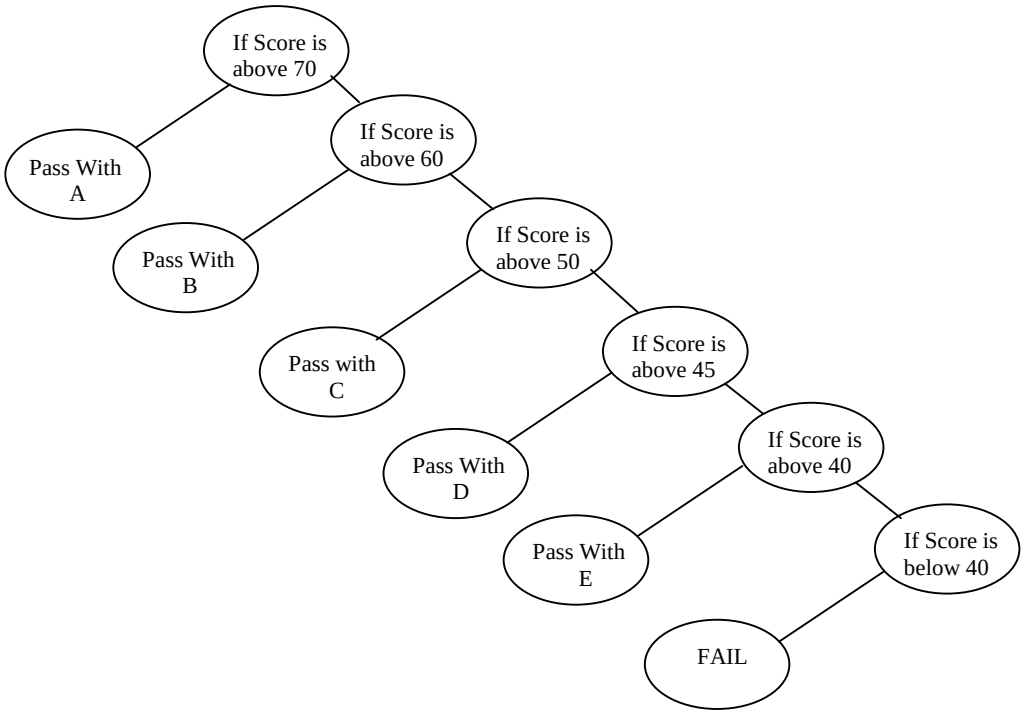


Fig 3: A Tree Representation of the Student Failure GP

3.3.2 Multiple Gene Symbolic Regression Genetic Programming (GP) Model:

A multigene individual consists of one or more genes, each of which is a "traditional" GP tree. Genes acquired incrementally by individuals in order to improve fitness (e.g. to reduce a model's sum of squared errors on a data set). The overall model is a weighted linear combination of each gene. Optimal weights for the genes are automatically obtained (using ordinary least squares to regress the genes against the output data). The resulting pseudo-linear model can capture non-linear behaviour. In multigene symbolic regression based GP, each prediction of the output variable y is formed by the weighted output of each of the trees/genes in the multigene individual plus a bias term. Each tree is a function of zero or more of the N input variables x1, … xn.

Mathematically, a multigene regression model can be written as:

$$y = d_o + d_1*tree1+...+d_m*tree_M ..... (1)$$

where do = bias (offset) term and d1,..., dm are gene weights and M is the number of genes (i.e. trees) comprising the current individual. The weights (i.e. regression coefficients) are automatically determined by a least squares procedure for each multigene individual.

The number and structure of the trees is evolved automatically during a run (subject to user defined constraints) using training data, i.e. a set of existing input values and corresponding output values. Testing data (another set of input and corresponding output values from the process or system you are modelling) can be used, after the run, to evaluate the evolved models. The testing data is not used to evolve the models and serves to give an indication of how well the models generalise to new data.
A pseudolinear multigene model of predictor output y with inputs x_1 to x_6; the weights d_0, d_1, d_2 are automatically obtained by least squares.

Multiple gene model (transfer function):

$$Y = d_o + d_1*x1+ d_2*x2+d_3*x3+d_4*x4+ d_5*x5+ d_6*x6 ……. (2)$$

The architecture of the developed system is given in Fig 4.

3.3.3 Solution Representation and Methodology

Genetic programming creates computer programs in the Lisp or scheme computer languages as the solution. Genetic algorithms create a string of numbers that represent the solution. In our model, an inductive-evolutionary approach is employed. The developed application based on Genetic programming uses the basic executional steps as follows to solve the failure rate problem:

- Step 1: Input data file containing RxC matrix of input (predictor) variables, an Rx1 matrix of output (response) variables, and system configuration file is fed to the GP system
- Step 2: Generate a random initial population of expressions based on Step1 using a multi-gene model
- Step 3: Assign a fitness value to each individual in the population
- Step 4: Create a new population of model expressions
- Step 5: Choose the best existing solutions
- Step 6: Create new solutions by crossover and mutation
- Step 7: The best solution is chosen as the fittest solution (least error margin).
- Step 8: Based on the best solution expression apply grammar based classification to establish the number of students below failure value
- Step9: Compute failure rate and print solution or candidate failure rate expression

Fig 4 shows the architecture of a typical Multi-Gene GP sequence. It must be emphasized here that Step1 specifically builds the structure of the GP system. Step 2 to 7 is achieved using the rungp functional class [11].

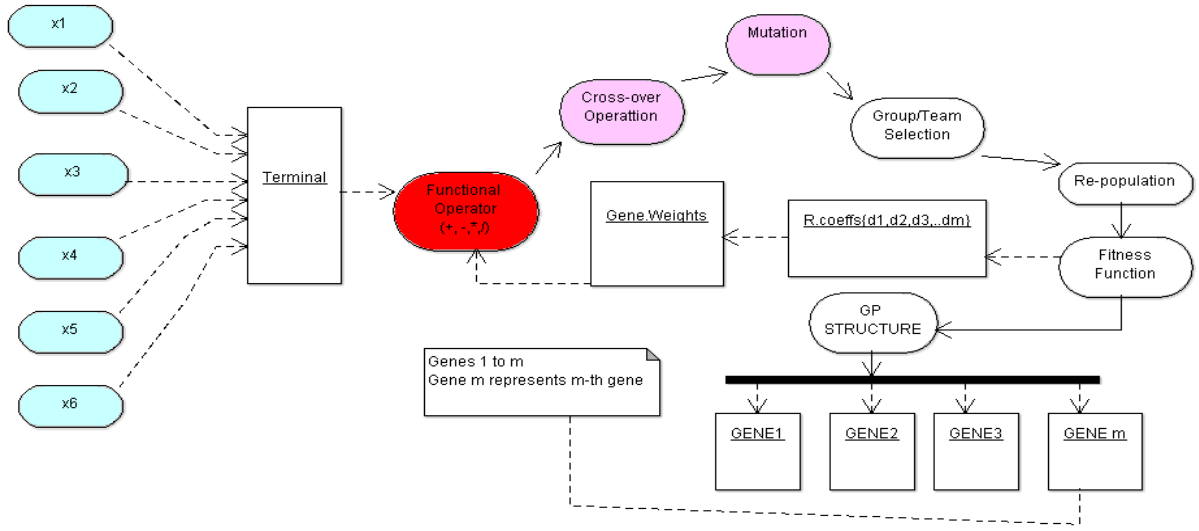


Fig 4: Architecture of a typical Multi-Gene GP Operation

3.3.4 Application Development

The application has been built in phases. The development phases use the concepts of windowing and navigation common to most modern GUI's. Windows are basically clickable user interfaces that facilitate data entry, retrieval and visualization in a compact and easy to use way using components such as buttons, text fields, menus, etc. The application is designed and developed in phases using callback functions or classes within a scripting program. It basically consists of three core windows (not shown):

- Start window: serves as front end for application
- Main window: for specifying core GP parameters and running the GP prediction system
- Grammar Window: for failure rate classification and estimation. A display has been added to the Main window and Grammar window to print the solution model expression for prediction and failure classifications respectively. It is possible for the user to move between windows by using callback buttons or menus. Fig 5 shows the class diagram including the core computational functions used in the developed application.

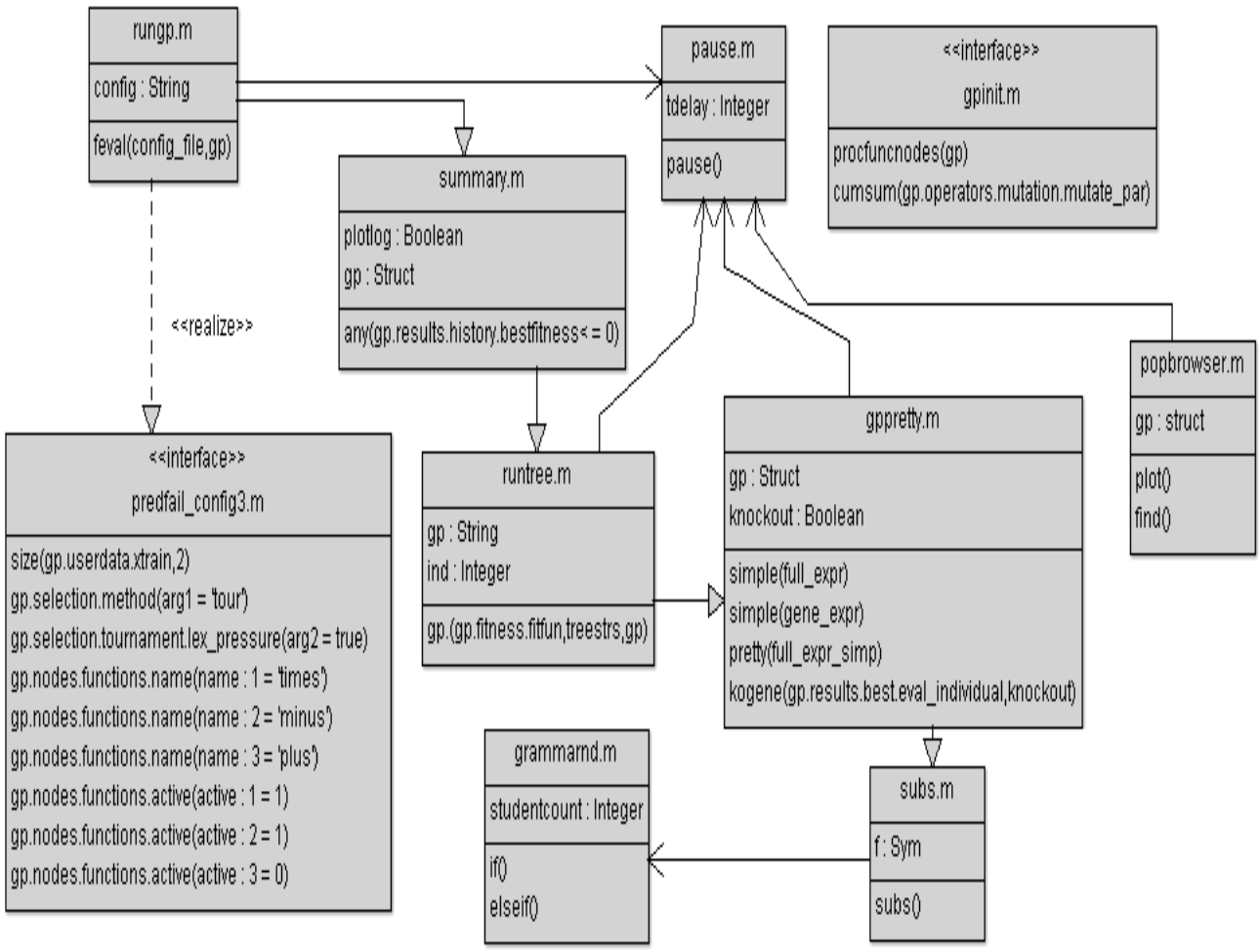


Fig 5: Computational Class Diagram of GPSFARPS

3.3.2 GP Model Input Data and Specifications

A continuous set of student raw scores used in the developed GP applications are proposed with sample student net scores labelled Q1 to Q5, and net continuous assessment, labelled CA score, jointly referred to as the Predictor Variables, while the output parameter (response variable) is the total score (TOTAL) which is as usual is graded and scaled to a maximum of 100 and minimum of 0. These data were obtained from historical records and are provided in Table 1. We have made the assumption that the scores represent a time series so future forecasts can be easily achieved and assured even when there is great variation in size, the principles remain the same i.e. we could easily scale down to a suitable range. In Table2, the GP parameters for the analysis and prediction process are presented taking into account bloat constraints [8].

Table1: Sample of Student Raw Score Data

Time	x1	x2	x3	x4	x5	EXAM SCORE	C.A. SCORE (x6)	TOTAL	GRADE
1	0	7	0	0	0	7	16	23	F
2	6	7	0	6	5	24	20	44	E
3	1	6	3	0	0	10	16	26	F
4	0	4	4	3	2	13	17	30	F
5	0	4	2	2	1	9	16	25	F
6	0	5	0	9	6	20	20	40	E
7	0	2	1	1	0	4	16	20	F
8	0	4	0	0	0	4	16	20	F
9	5	10	0	0	5	20	20	40	E
10	0	11	4	6	5	26	21	47	D
11	8	4	2	0	0	14	16	30	F
12	0	8	6	6	0	20	21	41	E
13	0	7	0	0	4	11	16	27	F
14	0	2	0	0	0	2	16	18	F

Table 2: GP Parameter Settings

GP algorithm parameters	Parameter Settings
Population size	150
Number of generations	500
Selection method	Plain Lexicographic Tournament Selection
Tournament size	4
Termination criteria	0
Maximum depth of each tree	4
Maximum number of trees	4
Maximum Number of Genes	4
Mathematical operators for symbolic regression	{+ , – , x}

4. RESULTS

In this study GPSFARPS application is used to train the raw score data from table 1 with specifications given in Table 2. A lexicographic tournament selection strategy is chosen for selecting the parent genes from the pool of available solutions. The tournament size is set to 4. The maximum depth of each tree in the multi-gene representation is set to 4 and this allows some control over the complexity of the evolved expressions. The instruction set or the functions that are used for symbolic regression are {+ , – , x}.

Figure 6 shows the final population of the GP run showing the trade-off between the accuracy of the fit and the complexity of the evolved models. The solutions points labelled in blue represent the set of dominated solutions while those in green represent the set of non-dominated solutions or the Pareto front. Solutions which are on the Pareto front are non-dominated in the sense that there are no solutions which have both a lower fitness and a lower model complexity simultaneously than these ones. In other words, if another solution has a lower fitness value then it must have a higher model complexity and vice-versa [12]. Fig 7 also captures the individuals involved in the best solution space.

The MGGP convergence characteristics in Figure 8 also indicate that, 30 generations is sufficient for the convergence of the algorithm. It can be seen that the objective function does not change significantly near the end of the GP run. Also the associated curve of the mean fitness is plotted below it. It shows that the overall population loses diversity very quickly and running the GP algorithm for more number of generations is not going to yield a much better solution.

However, it can also be seen that the algorithm should not be run for less than 10 generations for the present case, as the solution would not have converged sufficiently by then. The generated modelled expression is captured in Fig 9 which can be written as:

$$Y = x6 + x5 + x3 + x4 \quad \ldots\ldots\ldots\ldots\ldots (3)$$

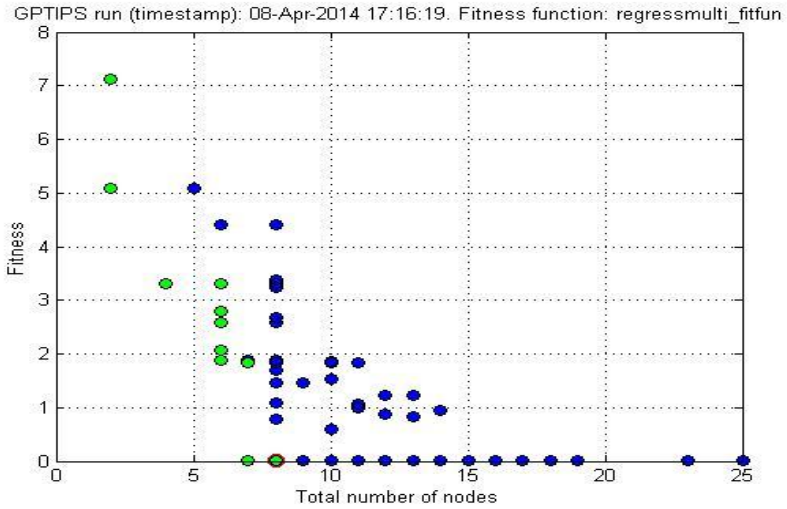


Fig 6: Fitness vs. complexity of the evolved Multi-Gene GP solutions along with Pareto solutions.

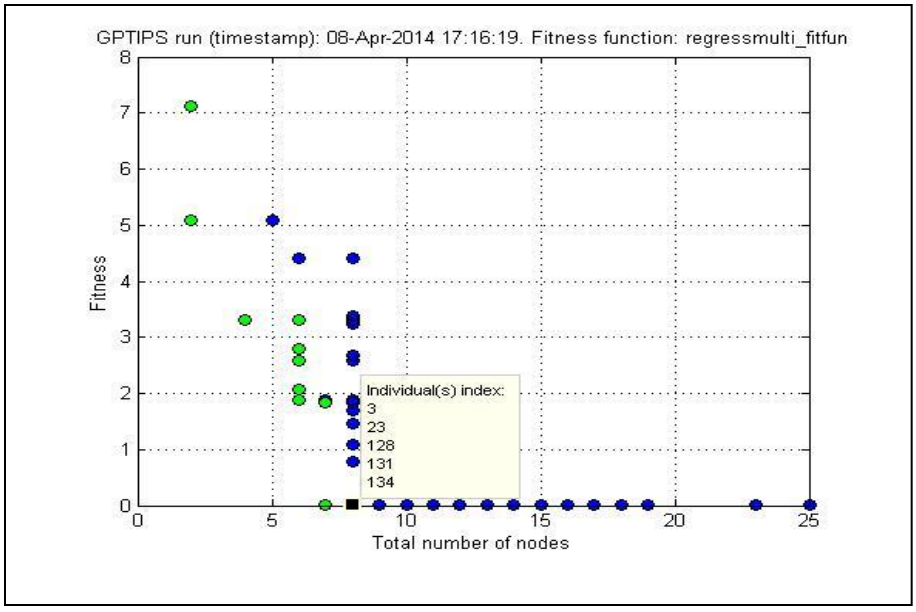


Fig 7: Fitness vs. complexity of the evolved Multi-Gene GP solutions along with Pareto solutions and best individuals (3, 23, 128, 131, 134).

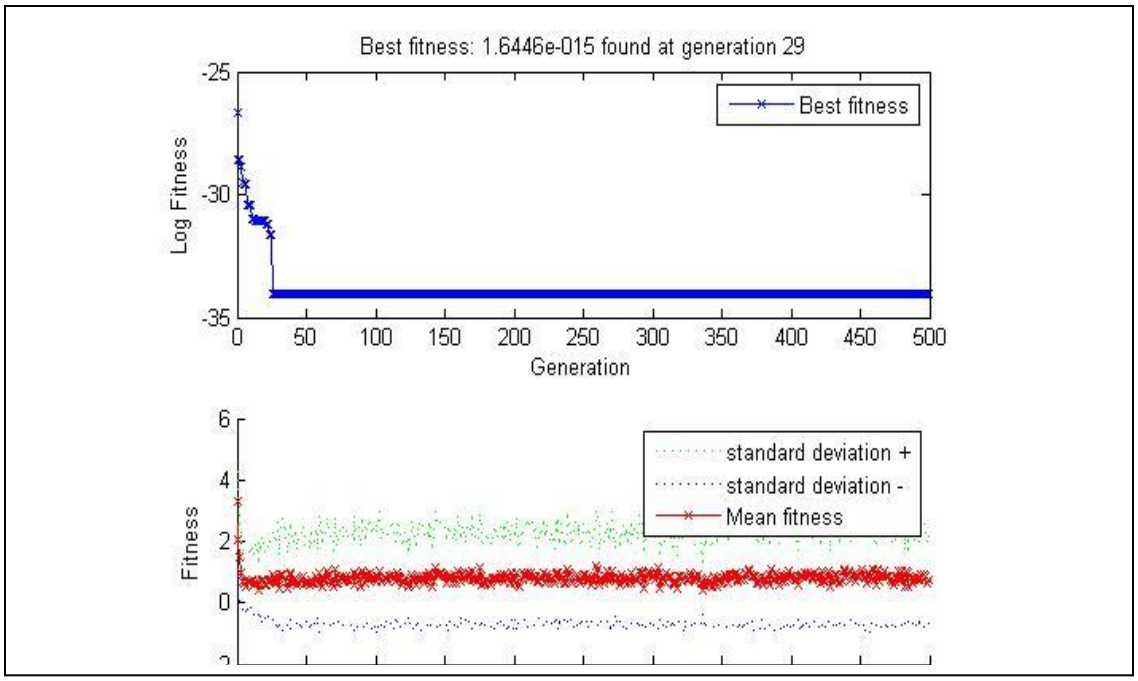


Fig 8: Convergence curves for the entire Multi-gene GP run with the best and mean fitness of solutions

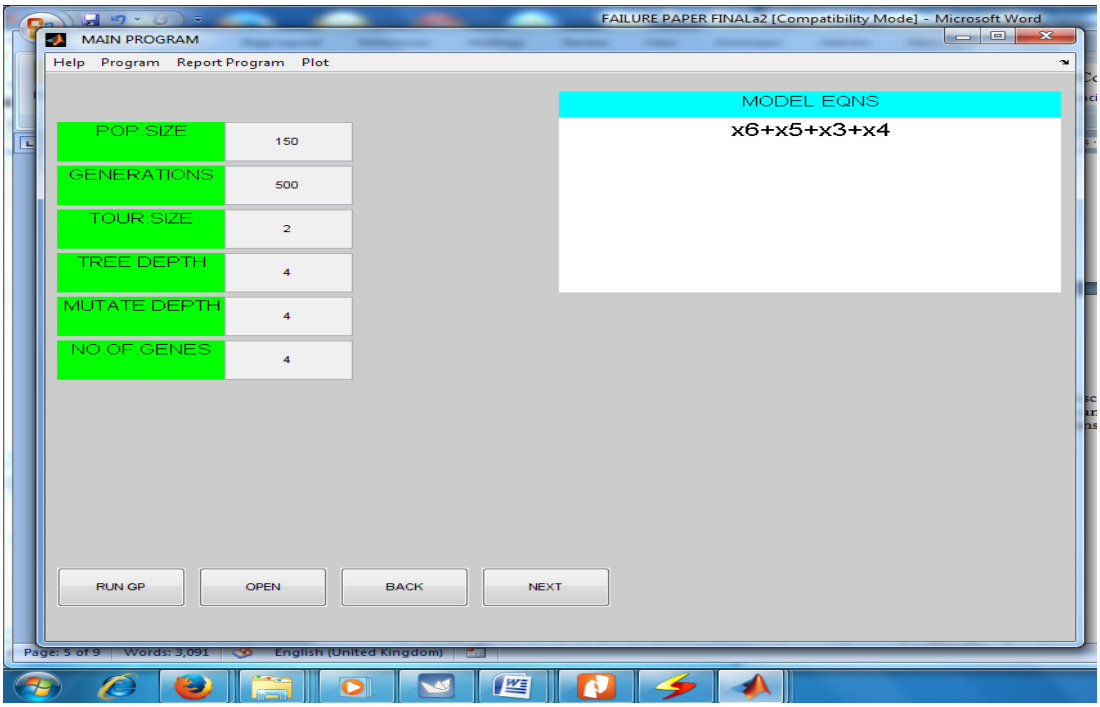


Fig 9: Snapshot of Generated model equation in GPSFARPS

5. DISCUSSIONS

In this study we have implemented the principles of GP using the multi-gene approach in a friendly user interface application. We applied the concept of cross-over and mutation to evolve a random population of individual solution programs and obtained a best fitted model for the specified GP parameters. The graphical reports are generated using MATLAB © handle graphics technology [15]. Using the interactive software GPSFARPS, GP parameters can be specified and meaningful features observed at run time.

A parsimonious model can be achieved if the core GP parameters, in particular, the number of generations, population size, tournament size, tree depth, number of gene trees and number of genes are constrained. From Fig 8, we have shown that specifying arbitrary number of generations does not significantly improve the system performance beyond possible achievable limits (Log Fitness versus Generation Plot). We should also point out here that the problem of local minima common to other AI schemes such as neural networks and fuzzy logic can be avoided with this evolutionary technique. Fast convergence has been achieved at approximately 30 generations from a maximum of 500 generations. This fast convergence rate may be attributed to the close correlation between predictor and response variables but should not be expected at all times. Also four candidate response variables, x6, x5, x4, and x3, were generated by the system and then used in the prediction process. This clearly shows the strength of multi-gene GP in minimizing or simplifying complex predictor equations.

Another issue what considering is the need for flexible specification of functional operators or function data set(s), since the function set plays a vital role in the structure of the evolved expression, such a technique can be valuable if the school administrator has first-hand knowledge of the shape of the failure rate. The failure shape or failure rate graph may be discovered using simple graph operations on test data from which an initial guess for suitable candidate functions may be arrived at.

Once candidate function(s) have been deduced, it can then be programmed into GPSFARPS using end-user programming call back functions. In this way, novel solutions to the failure rate problem may be discovered by the end-users themselves.

6. CONCLUSIONS

A object-oriented evolutionary software application for evolving model expressions that can predict student failure rates at school has been developed. The system was fed with incomplete raw score data and simulated response was satisfactory with a fast convergence rate – about 30 generations for a specified maximum generation of 500 and a minimal model expression consisting of only 4 predictor variables from a total of six predictor variables. The system can safely be used to facilitate predictions on complex data by using the software interface. The application will be open sourced and made available on the web to facilitate further development.

7. SIGNIFICANCE AND RECOMMENDATIONS FOR FUTURE WORK

This paper has presented a software application for predicting student failure rate at school. The software is capable of regenerating failure rate model expressions using a time –series dependent student's raw scores as input parameters and these model expressions can in turn be used by school administrators to determine student failure rate at school

The software can be made more useful if a bloat checking mechanism is incorporated. Further investigations need to be carried to test the effect of two state logic data on the GP software system developed, improving the function set database and making the program platform independent. Also, One-step in this direction will be to build a suitable software framework that will encourage end-user programming of the function data set. An interactive-agile function data set to facilitate end user programming of GPSFARPS in a future design is thus proposed.

REFERENCES

[1] Annelize J, 2009. "The Failure Rate of Hospitality Management Students in Win Courses at Tshwane University of Technology: Exploring Probable Causes. Faculty of Management Sciences Tshwane University of Technology, South-Africa

[2]] Sarah J.P, Overcoming Poor Failure Rates in Mathematics for Engineering Students: A Support Perspective. Harper Adams University College, Harper Adams Learning and Teaching Support Services, pp. 1-8

[3] Fire M, Katz G., Elovici Y, Shapira B, and Rokach L, 2012. Predicting Student Exam's Scores by Analyzing Social Network Data. AMT'12 Proceedings of the 8th international conference on Active Media Technology Springer-Verlag Berlin, Heidelberg, pp. 584-595, doi:10.1007/978-3-642-35236-2_59

[4] Vera C.M., Romero C., and Ventura S., 2011. Predicting Student Failure Using Data Mining, Journal of Applied Intelligence, , MA, USA, vol. 38 (8), Kluwer Academic Publishers Hingham, MA, USA, pp. 315-330.

[5] Johnson, William L., Johnson, Annabel M., and Johnson, Jared. Predicting Student Success on the Texas Chemistry STAAR Test: A Logistic Regression Analysis, http://eric.ed.gov/?id=ED534647, 2012

[6] Randall S. S, 2001. Neural Networks Refined: using a Genetic Algorithm to Identify Predictors of IS Student Success, The Journal of Computer Information Systems, vol. 41, Issue 3, pp. 42-47.

[7] Huang S., 2011. Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course. All Graduate Theses and Dissertations.Paper1086,http://digitalcommons.usu.edu/etd/1086.

[8] Huang S., and Fang, N., 2013. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models, Journal of Computers and Education, Elsevier, vol. 61, pp. 133-145.

[9] Kalles D., Pierrakeas C., 2006. Using Genetic Algorithms and Decision Trees for a posteriori Analysis and Evaluation of Tutoring Practices based on Student Failure Models, in IFIP Intemational Federation for Information Processing, vol. 204, Artificial Intelligence Applications and Innovations, Boston: Springer, pp. 9-18.

[10] Rquez-Vera C.M., Cano A., Romero C., and Ventura S., 2013. Predicting student failure at school using Genetic Programming and different data mining approaches with high dimensional and imbalanced data, Springer Science Business Media, vol. 38, pp. 315- 330, doi: 10.1007/s10489-012-0374-8.

[11] Searson P., 2009. GPTIPS: Genetic Programming & SymbolicRegressionforMATLAB,http://gptips.sourceforge.net.

[12] Indranil P, Daya S.P and Saptarshi D., 2013. Global solar irradiation prediction using a multi-gene genetic programming approach, Journal of Renewable and Sustainable Energy, vol.5, no.6, pp.1-30, doi: 10.1063/1.4850495.

[13] Riccardo P., William B.L., Nicholas F.M., and Koza J.R, 2008. A Field Guide to Genetic Programming, pp. 2

[14] Jesuína M., Fonseca B., and Joseph E. C., 2006. Secondary Student Perceptions of Factors affecting Failure in Science in Portugal, Eurasia Journal of Mathematics, Science and Technology Education, pp. 82-95

[15] Matlab version R2007b, 2007. Mathwork s Inc., USA, http://mathworks.com

[16] Romero C., Ventura S., Espejo P.G., and Hervás C., 2008. Data Mining Algorithms to Classify Students, Proceedings of the 1st International Conference on Educational Data Mining, Montréal, Québec, Canada, June 20-21, 2008

[17] Elite Research, 2013. Multiple Regression, Kean University, pp. 1-8

[18] Zlatko J.K., 2011. Predicting student success by mining enrolment data, Research in Higher Education Journal, pp.1-20.

[19] Chen, G. 2000. Discovering Decision Knowledge from Web Log Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube Technology, Journal of Educational Computing Research

[20] Yudelson, M.V., 2006. Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS, AAAI Workshop on Educational Data mining, pp. 1-8

[21] Cocea, M., and Weibelzahl, S., 2006. Can Log Files Analysis Estimate Learners' Level of Motivation?, Workshop on Adaptivity and User Modeling in Interactive Systems, Hidelsheim, pp. 32-35

[22] Hamalainen W. , Vinni, M., 2011. Classifiers for educational data mining., Chapman & Hall/VCRC London.pp 120- 340

[23] Fausett, L., Elwasif, W., 1994. Predicting Performance from Test Scores using Backpropagation and Counter propagation, IEEE Congress on Computational. pp. 112 - 320

[24] Martinez, D., 2001. Predicting Student Outcome using Discriminant Function Analysis, Annual Meeting of the Research and Planning Group, California, pp 163-173

[25] Superby, J.F., Vandamme, J.P., Meskens, N., 2006. Determination of Factors Influencing the Achievement of the First-year University Students using Data Mining Methods, Workshop on Educational Data Mining, pp. 37-44.

[26] Bravo, J., 2009. Checking the Reliability of GeSES: Method for Detecting Symptoms of Low Performance, Proceedings of the 9th International Conference on Intelligent Systems Design and Applications.
[27] Baker, J., 2002. Developing a generalizable detector of when student game the system, User Modeling and User-Adapted Interaction, April 02
[28] Tanner T., and Toivonen H., 2010. Predicting and preventing student failure – using the *k*-nearest neighbour method to predict student performance in an online course environment, Dept of Comp.Sc Helnsiki, Finland, pp.1-23

Authors' Brief

Mr. Orove, Osu Joshua obtained his B. Sc. degree in Computer Science and Statistics from University of Nigeria, Nsukka, Enugu state of Nigeria and he is currently concluding his M. Sc degree in Computer Science in the University of Port Harcourt, Nigeria. His interest is in Artificial Intelligence.

Dr. B.O. Eke is a Lecturer in the Department of Computer Science University of Port Harcourt, Nigeria. He obtained his B. Sc and M. Sc degrees in Computer Science from the University of Port Harcourt. His research interest is on working software as a signatory to the Agile manifesto.

N.E. Osegi obtained his HND in Electrical/Electronic Engineering from the Rivers State Polytechnic Bori Nigeria and is currently a postgraduate research scholar in the National Open University of Nigeria (NOUN). His specific research has focused on the use of software engineering tools, complex machine learning frameworks to design and implement physical and analytic computing systems, reasoning systems, control and automatics including robotics and artificial life systems. He can be reached by phone on +2347030081615 and through E-mail geeqwam@gmail.com, or nd@osegi.com

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Frequent Pattern and Association Rule Mining from Inventory Database Using Apriori Algorithm

K.S. Adewole, A.G. Akintola & A.R. Ajiboye
Department of Computer Science
University of Ilorin
Ilorin, Kwara State, Nigeria
adewole.ks@unilorin.edu.ng, abimbolar@yahoo.com, ajibraheem@unilorin.edu.ng

S.O. Abdulsalam
Department of Computer, Library and Info. Science
Kwara State University
Malete, Kwara State, Nigeria
sulepy@yahoo.com,

ABSTRACT

Recently, data mining has attracted a great deal of attention in the information industry and in a Society where data continue to grow on a daily basis. The availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge is the major focus of data mining. The information and knowledge obtained from large data can be used for applications ranging from market analysis, fraud detection, production control, customer retention, and science exploration. A record in such data typically consists of the transaction date and the items bought in the transaction. Successful organizations view such databases as important pieces of the marketing infrastructure. This paper considers the problem of mining association rules between items in a large database of sales transactions in order to understand customer-buying habits for the purpose of improving sales. Apriori algorithm was used for generating strong rules from inventory database. It was found that for a transactional database where many transaction items are repeated many times as a superset in that type of database, Apriori is suited for mining frequent itemsets. The algorithm was implemented using PHP, and MySQL database management system was used for storing the inventory data. The algorithm produces frequent itemsets completely and generates the accurate strong rules.

Keywords: Apriori Algorithm, data mining, database, strong rules & inventory.

African Journal of Computing & ICT Reference Format:
K.S. Adewole, A.G. Akintola & A.R. Ajiboye & S.O. Abdulsalam (2014). Frequent Pattern and Association Rule Mining from Inventory Database Using Apriori Algorithm. Afr J. of Comp & ICTs. Vol 7, No. 3. Pp35-42.

1. INTRODUCTION

Data mining is an interdisciplinary sub-field of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a dataset and transform it into an understandable structure for further use [1]. Advances in data collection and storage technology have led organizations to store vast amounts of data pertaining to their business activities. Extracting useful information from such huge data collections is of importance in many business decision making processes. Such an activity is referred to as Data Mining or Knowledge Discovery in Databases (KDD).

Data mining includes tasks such as Classification, Similarity Analysis, Summarization, Sequential Pattern Discovery, and Association Rule Mining among others [2]. Data mining or Knowledge discovery in databases (KDD) is the process of discovering previously unknown patterns form the huge amount of data stored in flat files, databases, data warehouses or any other type of information repository. Database mining deals with the data stored in database management systems such as MySQL [3].

The task of association rule mining is to find correlation relationships among different data attributes in a large set of data items, and this has gained lot of attention since its introduction. Such relationships observed between data attributes are called association rules [4]. A typical example of association rule mining is the market basket analysis. Consider a retail store that has a large collection of items to sell. Business decision regarding discount, cross-selling, grouping of items in different aisles, and so on needs to be made often in order to increase the sales and hence the profit. This inevitably requires knowledge about past transaction data that gives the buying habits of customers. The association rules in this case will be of the form "customers who bought item A also bought item B," and association rule mining is to extract such rules from the given transaction data history.

Market basket analysis, also known as association rule mining, is a method for discovering consumer purchasing patterns by extracting associations or co-occurrences from the stores' transaction database. The most well-known methodology called Apriori, was introduced by Agrawal et al. [4] which, as all algorithms for finding large itemsets, can be stated as follows. Given two non-overlapping subsets of product items, *X* and *Y*, an association rule in form of $X => Y$ indicates a purchase pattern that if a customer purchases *X* then he or she also purchases *Y*. Two measures, support and confidence, are commonly used to select the association rules [4].

In computer science and data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions such as collections of items bought by customers, or details of a website frequentation. Other algorithms are designed for finding association rules in data having no transactions or having no timestamps [2]. In this information age, due to the use of sophisticated technologies such as computers, satellites, etc., we have been collecting tremendous amounts of information like business transactions, scientific data, medical data, satellite data, surveillance video & pictures, world wide web repositories to name a few. With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, if not necessary, to develop powerful means for analysis and perhaps an interpretation of such data and for the extraction of interesting knowledge that could help in decision-making.

The kinds of patterns that can be discovered depend upon the data mining tasks employed. By and large, there are two types of data mining tasks: descriptive data mining tasks that describe the general properties of the existing data, and predictive data mining tasks that attempt to make predictions based on inference on available data. One of the popular descriptive data mining techniques is Association Rule Mining (ARM), owing to its extensive use in marketing and retail communities in addition to many other diverse fields. Mining association rules is particularly useful for discovering relationships among items from large databases [2].

Association rule mining deals with market basket database analysis for finding frequent itemsets and generate valid and important rules. Various association rules mining algorithms have been proposed by Agrawal et al. [4]. These algorithms include Apriori, Apriori-TID and Apriori Hybrid. Other algorithms for finding frequent itemsets include pincer search [3], FP (frequent pattern) tree [5]. A frequent pattern tree generates frequent itemsets without candidate generation.

1.1 Market Basket Analysis

Market basket analyses are an important component of analytical system in retail organizations. There are several definitions of market basket analysis. In a broader meaning, market basket analysis targets customer baskets in order to monitor buying patterns and improve customer satisfaction [6]. The following analytics can be used: attachment rates, demographic baskets, brand switching, customer loyalty, core items, items per basket, in-basket price, revenue contribution, shopper penetration and others.

In a narrower meaning, market basket analysis gives us the answer to the following question: which goods are sold together within the same transaction or to the same customer? By analyzing this information, we try to find out persistent patterns in order to offer related goods together and therefore increase the sales. We can track related sales on different levels of goods classifications or on different customer segments. Market basket analysis can also help retailers plan which items to put on sale at reduced prices [2]. Figure 1 shows market basket analysis.

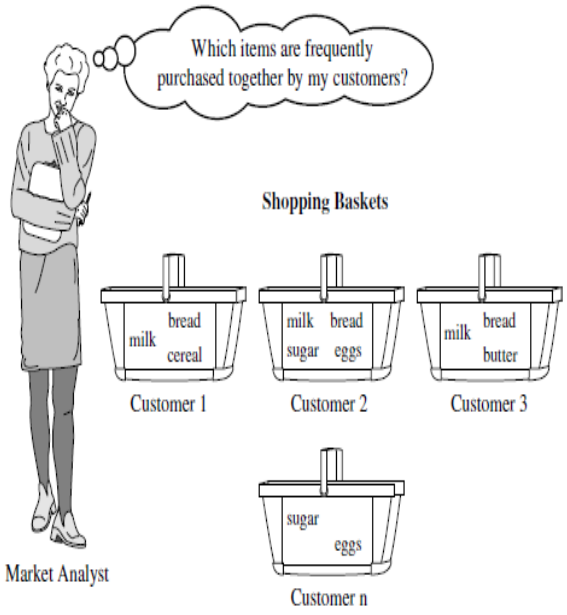


Figure 1: Market basket analysis adopted from [2]

The remaining part of this paper is organized as follows. Section 2 discusses the relevant related research carried out by the previous researchers in association rule mining. Section 3 focuses on methodology with main emphases on the Apriori algorithm used in this paper. Section 4 discusses the results obtained while Section 5 concludes the research work.

2. RELATED WORK

Data mining in education using association rule mining technique for identifying students' failure patterns was proposed by [7]. This approach identifies hidden relationship between the failed courses and suggests relevant causes of the failure to improve the low capacity students' performances. The researchers observed that the execution time of the proposed approach is inversely proportional to minimum support, since it increases as minimum support decreases, which confirmed increase in system complexity and response time as the minimum support decreases. It was also observed that to have a less complex system and a constructive, interesting and relevant patterns, the minimum confidence and support should be large enough to trash out coincidence patterns. In their research, 19 frequent itemsets and 114 rules were generated.

The researchers concluded that all the rules with confidence 1, are very strong rules, which implies that if a student failed the determinant (antecedent) course(s), such student will surely fail the dependent (consequent) course(s). It was recommended that this rule should be placed in curriculum structure. Also, if the rule support is higher, it means that all the courses involved are failed together by most of the considered students. Their proposed approach assists in the curriculum structure and modification in order to improve students' academic performance and trim down failure rate.

Bin, Wynne and Yiming [8] proposed hybridized mining algorithms for rule generation and for classification. The proposed algorithm is called CBA (Classification Based on Associations) which consists of two parts; a rule generator (CBA-RG) and a classifier builder (CBA-CB). The rule generator is based on Apriori algorithm for finding association rules. The CBA-RG generates set of class association rules (CARs) which consists of all the possible rules (PRs) that are both frequent and accurate. This is done by making multiple passes over the data. The proposed CBA-CB algorithm uses CARs to produce the best classifier out of the whole set of rules. This involves evaluating all the possible subsets of the training data and selecting the subset with the right rule sequence that gives the least number of errors. The researchers concluded that the classifier built this way is more accurate than that produced by the state-of-the-art classification system like C4.5. Furthermore, this integration helps to solve a number of problems that exist in the current classification systems.

Pragya, Madan, and Nupur [9] carried out a study on Apriori algorithm. An association driven mining application developed in Java was proposed. This application can be used to manage retail businesses that provide retailers with reports regarding prediction of product sales trends and customer behavior. The proposed approach improved Apriori algorithm in terms of running time, number of database scan, memory consumption and the interestingness of the rules over the classical Apriori Algorithm. The researchers proposed a mining algorithm for incremental large itemsets generation. Frequent itemsets discovered depends on value of parameters like support and number of transactions read at a time. Thus, execution time of the algorithm depends on transactional datagroups and minimum support value.

Margaret, Yongqiao, Le and Zahid [10] conducted an extensive survey of association rule mining. In their research, previous studies on association rule mining were explored and a brief classification strategy and performance comparison was carried out. A Genetic Algorithm for mining quantitative association rules was proposed by [11]. This algorithm was named QuantMiner. QuantMiner is a genetic-based algorithm for mining quantitative association rules. In QuantMiner, an item is either an expression $A = v$, where A is a categorical (also called qualitative) attribute and v is a value from its domain, or an expression $A \in [l, u]$ where A is a quantitative attribute.

According to the researchers, QuantMiner works directly on a set of rule templates. A rule template is a preset format of a quantitative association rule, either chosen by the user or computed by the system. For each rule template, the algorithm looks for the best intervals for the numeric attributes occurring in that template, using a Genetic Algorithm. Figure 2 shows the proposed QuantMiner algorithm together with its fitness function.

Function Fitness ($A \Rightarrow B$)
TempFitness = Gain($A \Rightarrow B$)
if *TempFitness* ≥ 0 **then**
  foreach *interval I in* $A \Rightarrow B$ **do**
    //favor small intervals
    TempFitness * = (1-Prop(I))^2
    if *Support*$(A \Rightarrow B) < MinSupp$ **then**
      //penalize low support rules
      TempFitness - = Nbtuples
return TempFitness

Algorithm 2: QUANTMINER
Input: A dataset composed of NbTuples, PopSize, GenNb, CR, MR, MinSupp, MinConf
Output: Quantitative association rules $\mathcal{R}$
Select a set of attributes
Let $\mathcal{R}_t$ a set of rule templates defined on these attributes
Compute the set of frequent itemsets on categorical attributes in $\mathcal{R}_t$
$\mathcal{R} = \emptyset$
foreach $r \in \mathcal{R}_t$ **do**
  Generate a random population POP of PopSize instantiated rules following the template r
  i=1
  while $i \leq GenNb$ **do**
    Form the next generation of population by mutation and crossover w.r.t. MR and CR.
    Keep PopSize rules in POP with the best Fitness values
    i++
  $\mathcal{R} = \mathcal{R} \cup Argmax_{R \in POP} Fitness(R)$
return $\mathcal{R}$

Figure 2: QuantMiner algorithm proposed by [11]

The researchers concluded that QuantMiner is very useful as an interactive data mining system that provides a better optimization criterion based on both support and confidence for keeping only high quality and interested rules.

3. METHODOLOGY

3.1 Association Rule Mining

Association Rule: An association rule is an implication of the form $X \Rightarrow Y$ [8], where *X* and *Y* are sets of items called itemsets and $X \cap Y = \emptyset$. Here, X is called the antecedent and Y is consequent. There are two important measures for association rules, these are support (*s*) and confidence ($\propto$) [10]. An itemset that contains ***k*** items is a ***k***-itemset.
The set {bread, butter, nylon} is a **3**-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known as the frequency, support count, or count of the itemset [2].

Support: The support (*s*) of an association rule is the ratio of the records that contain $X \cup Y$ to the total number of records in the database. Therefore, support ($X \Rightarrow Y$) = $P(X \cup Y)$. If a rule has 6% support value, this means that 6% of the total records contain $X \cup Y$.

Confidence: The confidence ($\propto$) of a given number of records is the ration of the number of records that contain $X \cup Y$ to the number of records that contain X. Therefore, confidence ($X \Rightarrow Y$) = $P(Y|X)$. Mathematically we have:

Confidence ($X \Rightarrow Y$) =

$$P(Y|X) = \frac{support(X \cup Y)}{support(X)} = \frac{support\_count(X \cup Y)}{support\ count(X)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

For example, if a rule has a confidence of 80%, this means that 80% of the records containing *X* also contain *Y*.

3.2 Apriori Algorithm

Apriori algorithm was proposed by Agrawal and Srikant 1994 as an algorithm for mining frequent itemsets for Boolean association rules. This algorithm works on the prior knowledge of frequent itemset properties such that k-itemsets are used to explore *(k + 1)*-itemsets. The set of frequent k-itemsets where $k \geq 1$ is found by scanning the database to accumulate the count for each item and later collects those items that satisfy the minimum support which is denoted by L_1. The set L_1 is then used to find L_2 which is the set of frequent 2-itemsets. This process continues until no more frequent k-itemsets can be found. According to [2], Apriori property is used to improve the efficiency of the level-wise generation of frequent itemsets. This property states that "*all nonempty subsets of a frequent itemset must also be frequent*". For instance, if an itemset *I* is not frequent, that is, it does not satisfy the minimum support threshold, then any other resulting itemset composed by adding an item say *O* to itemset I (i.e $I \cup O$) cannot occur more frequently than *I*.

This means that $P(I \cup O) < \min\_sup$ where *min_sup* is the minimum support threshold. Apriori algorithm follows a two-step process called join and prune. Below is the detail of the Apriori algorithm adapted from [2]. This algorithm was implemented in this paper using PHP and MySQL to discover frequent itemsets for mining Boolean association rules.

Procedure Apriori_Algorithm()
Input: *D*, a database of transactions; *min_sup*, the minimum support count threshold
Output: *L*, frequent itemsets in *D*

Steps:
(1) L_1 = find_frequent_1-itemsets(D);
(2) **for** (k = 2; $L_{k-1} \neq \phi$; k++) f
(3) C_k = apriori_gen(L_{k-1});
(4) **for each** transaction $t \in D$ { // scan D for counts
(5) C_t = subset(C_k, t); // get the subsets of t that are candidates
(6) for each candidate $c \in C_t$
(7) c.count++;
(8) **endfor**
(9) $L_k = \{ c \in C_k \mid c.count \geq min\_sup \}$
(10) **endfor**
(11) **return** $L = \cup_k L_k$;

Procedure apriori_gen(L_{k-1}: frequent (k-1)-itemsets)
(1) **for each** itemset $l_1 \in L_{k-1}$
(2) **for each** itemset $l_2 \in L_{k-1}$
(3) **if** ($l_1[1] = l_2[1]$)^($l_1[2] = l_2[2]$)^:::^($l_1[k-2] = l_2[k-2]$)^($l_1[k-1] < l_2[k-1]$) then {
(4) $c = l1 \times l2$; // join step: generate candidates
(5) **if has_infrequent_subset**(c, Lk-1) then
(6) **delete** c; // prune step: remove unfruitful candidate
(7) else add c to C_k;
(8) **endif**
(9) **return** C_k;

Procedure has_infrequent_subset(c: candidate k-itemset; Lk-1: frequent (k-1)-itemsets); //use prior knowledge
(1) **for each** (k-1)-subset s of c
(2) **if** $s \notin L_{k-1}$ then
(3) **return** TRUE;
(4) **return** FALSE;

4. RESULTS AND DISCUSSION

Figure 3 shows the interface used for setting minimum support and confidence threshold for the association rules generation. Item to be bought is typed inside the text box and the system searches the database to suggest items. The item is added to the transactional data by clicking on "Make Transaction" button.

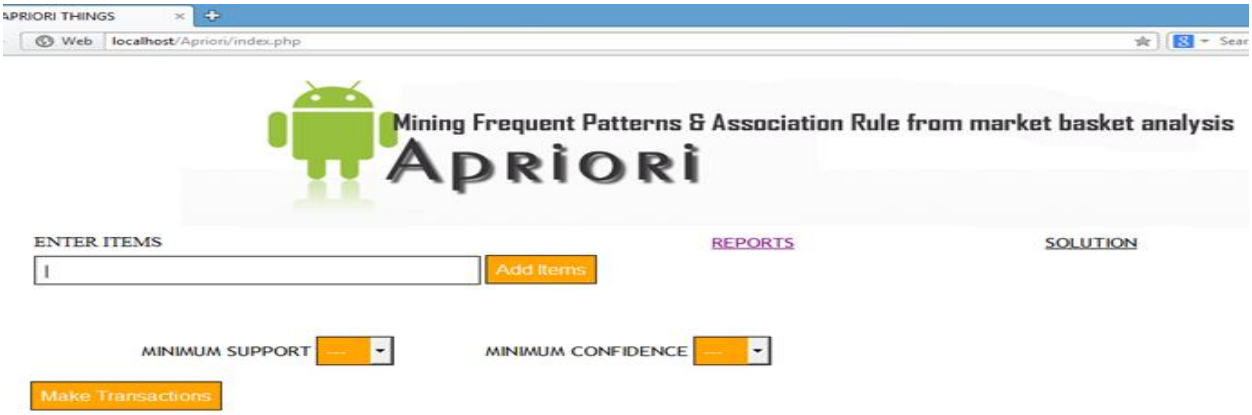


Figure 3: Interface for setting min_support and min_confidence

The solution menu is used to perform the join and prune steps of the Apriori algorithm as shown in Figure 4. The initial stage shows the transactional data used for the association mining followed by the frequent itemsets generation stages. Three itemsets were generated from the initial transactional data. These represent those itemsets, which satisfied the minimum support count threshold.

APRIORI during iterations (<<BACK) REPORTS STRONG RULE

ID	ITEMS
1	Camera WashingMachine GassCooker
2	DVDPlayer ElectricKettle Refrigerator
3	GassCooker Freezer Iron
4	TV SamsungS4 iPhone
5	TV WashingMachine Refrigerator
6	TV Generator TVStand
7	TV Generator TVStand
8	TV Generator TVStand
9	TV WashingMachine Refrigerator
10	GassCooker Freezer Iron
11	CDPack ExtensionBox
12	Camera WashingMachine GassCooker
13	Camera WashingMachine GassCooker

ID	ITEMS	SUPPORT COUNT
1	Camera	3
2	WashingMachine	5
3	GassCooker	5
4	DVDPlayer	1
5	ElectricKettle	1
6	Refrigerator	3
7	Freezer	2
8	Iron	2
9	TV	6
10	SamsungS4	1
11	iPhone	1
12	Generator	3
13	TVStand	3
14	CDPack	1
15	ExtensionBox	1

ID	ITEMS	SUPPORT COUNT
1	GassCooker	5
2	WashingMachine	5
3	Refrigerator	3
4	TV	6
5	Generator	3
6	TVStand	3

ID	ITEMS	SUPPORT COUNT
1	TV Generator	3
2	TV TVStand	3
3	Generator TVStand	3

Figure 4: Apriori algorithm during iterations

Figure 5 shows the strong rules generated after the computation of confidence on the three most frequent itemsets as satisfied by the threshold. This shows that if a customer purchases TV, there is 70% confidence that he/she will purchase TV Stand and Generator. Conversely, there is 85% confidence that a customer who purchases Generator will purchase both TV and TV Stand as well.

BACK REPORTS

STRONG RULES	
RULES	CONFIDENCE
TV --> Generator TVStand	70%
Generator --> TV TVStand	85%

Figure 5: Generated strong rules

5. CONCLUSION

This paper focuses on Apriori implementation for association rule mining. Frequent itemsets generated by Apriori algorithm completely depend on a minimum support threshold. It was observed that Apriori wastes running time due to candidate itemsets generation, hence, the need for a more robust hybrid algorithm for association rule mining is inevitable. However, for a transactional database where many transaction items are repeated many times as a superset in that type of database, Apriori is suited for mining frequent itemsets. Thus, this algorithm produces frequent itemsets completely. The proposed system can be used to determine the buying patterns of customers with greater rate of accuracy and subsequently improve daily sales.

REFERENCES

[1] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence. Retrieved 20th July, 2013 from http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf

[2] Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques, Morgan Kaufmann. USA: Elsevier Inc.

[3] Lin, D., Kedem, Z. M. (2002). Pincer-Search: An Efficient Algorithm for Discovering the Maximum Frequent Set, IEEE Transaction on Knowledge and Data Engineering, 14(3), pp. 553-566.

[4] Agrawal, R. Imielinski, T. and Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases, Proceedings of the ACM SIGMOD International Conference on Management of Data.

[5] Han, J., Pei, J., and Yin, Y. (2000). Mining Frequent Patterns without Candidate Generation, Proc. ACM SIGMOD.

[6] Microstrategy (2003). Business Intelligence in the Retail Industry, Microstrategy World 2003 Conference, Las Vegas.

[7] Oladipupo, O. O. and Oyelade, O. J. (2010). Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach, International Journal of Computer Science & Security (IJCSS), 4(2), pp.199-207.

[8] Bing, L., Wynne, H. and Yiming, M. (1998). Integrating Classification and Association Rule Mining, American Association for Artificial Intelligence. Retrieved 20th June, 2013 from http://www.cs.rit.edu/~dmrg/dm_winter/reading/integ_class_arm.pdf.

[9] Pragya, A., Madan, L. Y. and Nupur, A. (2013). Study on Apriori Algorithm and Its application in Grocery Store, International Journal of Computer Applications, 74(14), pp. 1-8. Retrieved 14th July, 2013 from http://www.ijcaonline.org.

[10] Margaret, H. D., Yongqiao, X., Le, G. and Zahid, H. (2009). A Survey of Association Rules. Retrieved 14th July, 2013 from http://www.coursehero.com/file/2789589/grue-assoc

[11] Ansaf, S., Christel, V., and Cyril, N. (2007). QuantMiner: A Genetic Algorithm for Mining Quantitative Association Rules. Retrieved 20th June, 2013 from http://www.eecs.wsu.edu

Authors' Biography

Adewole Kayode Sakariyah is a lecturer of at the Department of Computer Science, University of Ilorin, Ilorin, Kwara State, Nigeria. He obtained a National Diploma in Computer Science with distinction at the Federal Polytechnic, Ede, Osun State in 2005, a B.Sc Computer Science with first class honours at the University of Ilorin, Ilorin, Nigeria in 2009, a Master of Science Degree in Computer Science at the University of Ilorin, Ilorin, Nigeria. His research interests include Information Security, Artificial Intelligence, Data Mining and Mobile Computing. He is highly proficient in application software development using a wide range of programming languages. He can be reached by phone on +2348038544348 and through E-mail adewole.ks@unilorin.edu.ng

Akintola Abimbola Ganiyat is a Lecturer in the Department of Computer Science, Faculty of Communication and Information Sciences. University of Ilorin. Ilorin. She obtained a Bachelor degree (B.Sc.) in Computer Science from the Bowen University Iwo, Osun State, Nigeria in 2007, where she graduated with a Second Class Honours (Upper Division). A Master degree (M.Sc.) in Computer Science from the University of Ilorin in 2011. She is currently pursuing a doctoral degree (Ph.D) in Computer Science at University of Ilorin. Her research interests include Biometrics, Data Mining and Software Engineering. She can be reached through abimbolar@yahoo.com

Abdulsalam Sulaiman Olaniyi holds a Bachelor of Science (B.Sc.) Degree and Master of Science (Degree) in Computer Science both from University of Ilorin, Ilorin, Nigeria. He is currently a lecturer in the Department of Computer, Library and Information Science, Kwara State University, Malete, Nigeria. He is a member of Nigeria Computer Society (MNCS). His research areas include Data Mining, Software Engineering and Artificial Intelligence. He can be reached by phone on +2348030785287 and through E-mail sulepy@yahoo.com

Ajiboye Adeleke Raheem obtained B.Sc.(Computer Science) from the Department of Computer Science, University of Ilorin, Ilorin, Kwara State, Nigeria. He obtained a Master of Science in Computer Science at the same department in 2010. He joined the department as an Assistant Lecturer in 2011. In 2013, he commenced a PhD degree in Computer Science at the Universiti Malaysia Pahang. He is a member of Nigeria Computer Society (NCS) and Computer Professionals Registration Council of Nigeria (CPN). His research interests include: Data Mining, Computational Intelligence and Image Processing. Ajiboye has five publications to his credit in reputable International Journals. He can be contacted through e-mails: ajibraheem@unilorin.edu.ng ; ajibraheem@live.com

Vol 7. No. 3 - September, 2014

African Journal of Computing & ICT

www.ajocict.net

Necessary Conditions for Controllable Systems to be Observable

S.E. Aniaku
Department of Mathematics
University of Nigeria, Nsukka, Nigeria.
Phone: +2348065751038
E-mail: Stephen.aniaku@yahoo.com

ABSTRACT

Generally, the controllability and observability of controllable systems were treated separately. In this paper, through the use of duality condition, the necessary condition for controllable systems to be observable is derived.

Keywords: Controllability, Observability, Duality, Skew-Symmetric matrix, Transpose.

African Journal of Computing & ICT Reference Format:
S.E. Aniaku (2014). Necessary Conditions for Controllable Systems to be Observable. Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 43-46.

1. INTRODUCTION

Control system engineers are concerned with the problems of controllability, observability, and stability of dynamical systems. Observability is the ability to observe or to measure the output of all the parameters or state variables in the system. Controllability is the ability to move a system from any given state to another desired state, using the input u. The controllability condition is that the existence of such an input should be assured . Stability, on its own, is often phrased as the bounded response of the system to any bounded input. For any control system to be successful, it must have these properties; that is observability, controllability and stability properties. For linear control systems, such properties can be maintained with minimal conditions. For nonlinear control systems, uncertainties present a big challenge to the system engineers, who work hard to maintain these properties using limited information. For purpose of clarity, let us consider these properties one by one. Let us consider the time invariant system represented by:

$$\dot{x} = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) \qquad (1.1)$$

Where
x, u, y are n-vector, m-vector, and p-vector respectively and A, B, and C are respectively n x n, n x m, and p x n constant matrices.

2. CONTROLLABILITY.

For a linear system given by (1.1), if there exists an input u which transfers the initial state $x(0) = x_0$ to the state $x(t_1) = x_1$ in a finite time t_1, the state x_0 is said to be controllable.

If all the initial states are controllable, the system is said to be completely controllable. If $x_1 = \bar{0}$, the zero state, the system is said to be null-controllable Eke A.N.(2000). It is known, Stephen B. (1975) that the unique solution of the linear equation

$$\dot{x} = A(t)x(t) + B(t)u(t) \qquad (2.1)$$

$x(t_0) = x_0$
Is given by

$$x(t) = \varphi(t, t_0)\left[x_0 + \int_{t_0}^{t} \varphi(t, s)B(s)u(t)ds\right] \qquad (2.2)$$

Where $\varphi(t, s)$ is the transition matrix satisfying

$$\varphi(t, t_0) = A(t)\varphi(t, t_0) \text{ for } t \geq t_0$$

and $\varphi(0, 0) = I$, the identity matrix. The solution of (2.1) by Katsuhisa et al (1988), is given as

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}B(s)u(s)ds \qquad (2.3)$$

If the system is null-controllable, there exists an input $x(t_1) = x_1 = 0$ at a finite time $t = t_1$ such that multiplying (2.3) by e^{-At_1}, we get

$$-x_0 = \int_0^{t_1} e^{-As}B(s)u(s)ds \qquad (2.4)$$

So any controllable state will satisfy (2.4) and for a completely controllable systems every state x_0 in $\Re^n$ satisfies (2.4) with $t_1 > 0$. From (2.4) it is found that complete controllability of a system depends on A and B and is independent of the output matrix C.

Definition 2.1. Dauer J.P.(1971)

A set of vector functions $\{x_1, x_2, ..., x_k\}$ is said to be linearly independent over finite interval I if for every set of non zero vectors $(a_1, a_2, ..., a_k)\ \varepsilon\ C^k$, there exists a subset J of I with positive measures such that:

$$\sum_{i=1}^{k} a_i x_i \neq 0\ \forall\ t\ \varepsilon\ J\ .$$

The following simple theorems, due to Daeur J.P. (1971), characterize complete controllability; that is

Theorem 2.1

Assume $B\ \varepsilon\ L^{\infty}$, the system (1.1) is completely controllable in L^P if and only if the rows of the matrix functions $\{X^{-1}(t)B(t)\}$ are linearly independent over I.

Theorem 2.2

Let $2 \leq p < \infty$, and assume that $B\ \varepsilon\ L^P$, system (1.1) is completely controllable in L^P if and only if

$$W = \int X^{-1}(s)B(s)B^T(s)X^{-1^T}(s)ds$$

is positive definite.

Theorem 2.3

Suppose $A\ \varepsilon\ L^1$ and $B\ \varepsilon\ L^P$, $1 \leq p \leq \infty$ and assume that the system (2.1) is completely controllable in $L^q, \frac{1}{p}+\frac{1}{q}=1$. There exists $\varepsilon > 0$ such that if

$$|A-C|_P + |B-D|_P < \varepsilon$$

Then the system

$$y = C(t)y + D(t)u$$

is completely controllable in L^q.

Lemma 2.1 Dauer J.P (1971)1

Suppose W is positive definite n x n constant matrix. There exists $\varepsilon > 0$ such that if the constant n x n matrix V satisfies $|W-V| < \varepsilon$, then V is positive definite.

If we assume the conditions of Theorem 2.3, then, for every r $p \leq r < \infty$, there exists $\varepsilon > 0$ such that if

$$|A-C|_r + |B-D|_r < \varepsilon\ ,$$

Then (2.5) is completely controllable in L^q.

Apart from the above three rather simple theorems due to Dauer J.P. (1971), we have the following crucial fact in complete controllability, that is

Theorem 2.4 Katsuhisa et al (1988)

The necessary and sufficient condition for the system (1.1) to be completely controllable is one of the following equivalent conditions;

1. $W(0,t_1) = \int_0^{t_1} e^{-At} BB^T e^{-A^T t} dt$ is non singular

   [$W(0,t_1)$ is called controllability gramian and $(.)^T$ denotes matrix transpose]

2. The controllability matrix $\zeta = [B, AB, A^2B, ..., A^{n-1}B\ ]$ has rank n.

Proof:

Condition 1. Sufficiency:

If $W(0,t_1)$ given above is nonsingular, the following input can be applied to the system

$$u(t) = -B^T e^{-A^T t} W^{-1}(0,t_1)x_0$$

For the input (2.6), the state of the system (2.3) is given by

$$x(t_1) = e^{At_1}x_0 - e^{At_1}\{\int_0^{t_1} e^{-As} BB^T e^{-A^T s} ds\}W^{-1}(0,t_1)x_0 = 0$$

For any initial state x_0. Therefore, the system (A,B) is controllable.

Necessity:

Assume that although $W(0,t_1)$ is singular for any $t_1 > 0$, there exists a non-zero n-vector α such that

$$\alpha^T W(0,t_1)\alpha = \int_0^{t_1} \alpha^T e^{-As} BB^T e^{-A^T s}\alpha ds = 0 \quad (2.7)$$

which yields for any $t \geq 0$

$$\alpha^T e^{-At} B = 0^T \quad t \geq 0, \alpha \neq 0 \quad (2.8)$$

From the assumption of controllability, there exists an input u satisfying (2.4). Therefore from (2.4) and (2.8)

$$-\alpha^T x_0 = \alpha^T \int_0^{t_1} e^{-As} B(s)u(ds) = 0 \quad (2.9)$$

Holds for any initial state x_0

By choosing $x_0 = \alpha$, (2.9) gives $\alpha = 0$ which contradicts the non-zero property of α. Therefore the non-singularity of $W(0,t_1)$ is proved.

Condition 2. Sufficiency:

Let us first assume that if the rank of $\zeta = n,\ W(0,t_1)$ is singular. Therefore (2.8) that is $\alpha^T e^{-At} B = 0^T,\ t \geq 0,\ \alpha \neq 0$ holds. Derivatives of equation (2.8) at t=0 yields.

$$\alpha^T A^k B = 0^T. \quad \text{k = 0, 1, 2,…, n-1.} \quad (2.10)$$

Which is equivalent to

$$\alpha^T \left[B, AB, A^2B, ..., A^{n-1}B \right] = \alpha^T \zeta = 0^T \quad (2.11)$$

This contradicts the assumption that the rank $\zeta = n$, so the system is completely controllable.

Necessity:

Let us assume that the system is completely controllable, but rank $\zeta < n.$ From the Cayley- Hamilton theorem, Chi-Tsong.C (1984), A^{n+i} can be expressed as a linear combination of $I, A, ..., A^{n-1}$, which yields

$$\alpha^T e^{-At} B = 0^T, \quad t \geq 0, \alpha \neq 0.$$

So,

$$0 = \int_0^{t_1} \alpha^T e^{-As} BB^T e^{-A^T s}\alpha^T W(0,t_1)\alpha \quad (2.12)$$

Since the system is completely controllable, $W(0,t_1)$ is non singular from condition 1 above. Then α in (2.12) is zero, which contradicts the assumption that α is non-zero. So, rank $\xi = n,$ completing the proof of the theorem.

3. OBSERVABILITY.

When using the output of the system (1.1) measured from t=0 to time $t = t_1$, if the initial state $x(0) = x_0$ is uniquely determined, x_0 is said to be observable. The output of the system (1.1) is given by Katsuhisa et al (1998) as

$$y(t) = Ce^{At}x_0 + C\int_0^t e^{A(t-s)} Bu(s)ds \quad (3.1)$$

We note that the output and the input can be measured and used. So a signal η can be obtained from u and y using the formula, by Katsuhisa et al (1988)

$$\eta(t) = y(t) - C\int_0^t e^{A(t-s)} Bu(s)ds \quad (3.2)$$
$$= Ce^{At}x_0.$$

Since p is usually smaller than n, x_0 can not be determined uniquely by $\eta(t)$ is available over a time interval from 0 to t, and the system is completely observable, the initial state x_0 can be uniquely determined. If (3.2) is multiplied by $e^{A^T t}C^T$ and integrated from 0 to t_1, we get

$$\{\int_0^{t_1} e^{A^T t} C^T Ce^{At} dt\}x_0 = \int_0^{t_1} e^{A^T t} C^t \eta(t)dt \quad (3.3)$$

Let us define an n x n matrix $M(0,t_1)$ by;

$$M(0,t_1) = \int_0^{t_1} e^{A^T t} C^T Ce^{At} dt \quad (3.4)$$

If $M(0,t_1)$ is non singular, x_0 is determined uniquely from (3.3) as

$$x_0 = M^{-1}(0,t_1)\int_0^{t_1} e^{At}C\eta(t)dt \qquad (3.5)$$

From (3.5), we see that the non singularity of $M(0,t_1)$ for $t_1 \geq 0$ is a sufficient condition for the system (1.1) to be completely observable. $M(0,t_1)$ is non singular is proved in Katsuhisa F. et al. (1998) page 52.

This then leads to the following important theorem.

Theorem 3.1 Katsuhisa F. et al.(1998)
A necessary and sufficient condition for the system (1.1) to be completely observable is one of the following equivalent conditions:

1. $M(0,t_1) = \int_0^{t_1} e^{A^T t}C^T Ce^{At}dt$ is non singular.
2. The observability matrix defined as n x np matrix
$$[C^T, A^TC^T, (A^T)^2C^T, ..., (A^T)^{n-1}C^T]$$
has rank n.

Proof:
For 1, this has been proved from the derivation of (2.7). For 2, it can be proved from 1 in a similar way as in theorem 2.4.
We note that controllability and observability have a link between them as stated in the following duality theorem.

Theorem 3.2 (Duality) Stephen B (1975)
The system

$$\begin{aligned} x &= A(t)x(t) + B(t)u(t) \\ y &= C(t)x(t) \end{aligned} \qquad (3.6)$$

is completely controllable if and only if the dual system

$$\begin{aligned} x &= -A^T(t)x(t) + C^T(t)u(t) \\ y &= B^T(t)x(t) \end{aligned} \qquad (3.7)$$

is completely observable.

4. MAIN RESULT

The dual theorem leads us to the main result of this paper.
Theorem 4.1

The necessary condition for a controllable system

$$\begin{aligned} x &= A(t)x(t) + B(t)u(t) \\ y &= C(t)x(t) \end{aligned} \qquad (4.1)$$

to be observable is that the n x n matrix A be skew-symmetric and matrices B and C be transpose of each other.

Proof:
Since the system (4.1) is controllable, the controllability gramian $W(0,t_1)$ defined by

$$W(0,t_1) = \int_0^{t_1} e^{-At}BB^T e^{-A^T t}dt \qquad (4.2)$$

is non singular. From duality Theorem 3.2, (4.1) is observable if and only if the observability gramian $M(0,t_1)$ is defined by

$$M(0,t_1) = \int_0^{t_1} e^{A^T t}C^T C\, e^{At}dt \qquad (4.3)$$

must be non singular. By comparing (4.2) and (4.3), we see that if $A^T = -A$ and $C = B^T$, $(B = C^T)$ the two equations are the same. Since (4.2) is non singular, (4.3) will also be non singular, and so the obervability condition of (4.1) is assured. $A^T = -A$ implies that A is skew-symmetric, and $C = B^T$, $B = C^T$ means that they are of transpose of each other.

REFERENCES

[1] Chi-Tsong C. Linear system Theory and Design. 383 Madison Avenue. New York. (1984).
[2] Dauer J.P Perturbation of Linear Control Systems. SIAM J. Control Vol. 9 No.3 (1971)
[3] Eke A. N. Stabilizability For Linear Feedback Observable Systems. Journal of Maths. Society Vol. 19. pp59-68. (2000).
[4] Katsuhisa F. et al. State Variable Methods in Automatic Control. John Wiley & Sons. New York. (1988).
[5] Stephen B. Introduction to Mathematical Control Theory. Clarendom Press. Oxford (1975).

Author's Bio

Dr. S.E. Aniaku is a Lecturer 1 in the Department of Mathematics of University of Nigeria, Nsukka. He is a native of Aku in Igbo- Etiti Local Government Area of Enugu State.He obtained his N.C.E in 1980 from Alvan Ikoku Collage of Education, Owerri. He got his B. Sc. and M. Sc. in 1986 and 1998 respectively from University of Nigeria, Nsukka. He obtained his Ph.D degree in 2012 from Rivers State University of Science and Technology, Port-Harcourt. His research areas are in Control Theory and Modeling. He can be reached through phone number +2348065751038 and e- mail stephen.aniaku@yahoo.com.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Authenticating E - Banking Services in Nigeria through Digital Signatures

G. E. Okereke PhD
Computer Science Department
University of Nigeria, Nsukka
E-mail: george.okereke@unn.edu.ng
Phone Number +2348037784613

O. A. Ezugwu
Computer Science Department
University of Nigeria, Nsukka.
E-mail: assumpta.ezugwu@unn.edu.ng
Phone Number: +2348035689222

ABSTRACT

This paper examined in details the architecture of an electronic banking (E-banking) system. It described the meaning, actors and elements of electronic banking, its usefulness and problems. The paper went further to describe digital signature as a vital tool used to authenticate e-banking services. It provides guidance to financial institutions in Nigeria on identification and control of risks associated with electronic banking activities and discusses e-banking risks from the perspective of the services or products provided to customers.

Keywords: Electronic banking, Authentication, Authorization, Public Key Cryptography, and Digital Signature.

African Journal of Computing & ICT Reference Format:
G. E. Okereke & O. A. Ezugwu (2014). Authenticating E - Banking Services in Nigeria through Digital Signatures.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 47-52.

1. INTRODUCTION

Electronic banking also known as electronic fund transfer requires customer authentication. Reliable customer authentication is necessary for financial institutions engaging in any form of electronic transaction. A reliable authentication mechanism can help financial institutions minimize fraud and promote the legal enforceability of their electronic agreements and operations. Stronger customer authentication strategies are necessary to enforce anti-money laundering measures and help banks and other financial institutions detect and reduce identity theft. Customer interaction with financial institutions is migrating from physical recognition and paper-based operations to remote electronic access and transaction initiation.

The risks associated with business transactions with unauthorized or incorrectly identified individuals in an electronic transaction especially E banking environment could result in fund loss and reputation damage through fraud, disclosure of confidential information, corruption of data, unenforceable agreements among several others. There are a variety of authentication strategies financial institutions can adopt to authenticate customers.

These include the use of passwords and personal identification numbers (PINs), digital certificates using a public key infrastructure (PKI), physical devices such as smart cards or other types of "tokens," Kerberos, database comparisons, biometric identifiers and radio frequency identification (RFID) or other forms of “contactless” technology that scans customer information without direct contact between a customer and her bank. [1, 2] The level of risk protection afforded by each of these tools varies and is evolving as technology evolves.

Existing authentication strategies involve three basic "factors":

- What the user *knows* (e.g., password, PIN);
- What the user *possesses* (e.g., ATM card, smart card); and
- Who the user *is* (e.g., biometric characteristic, such as a fingerprint or retinal pattern).

Authentication methods that depend on more than one factor typically are more difficult to compromise than single factor systems. Accordingly, properly designed and implemented multi-factor authentication methods are more reliable indicators of authentication and stronger fraud deterrents in electronic banking.

For example, the use of a logon ID/password is single factor authentication (i.e., what the user knows); whereas, a transaction using an ATM typically requires two-factor authentication: something the user possesses (i.e., the ATM card) combined with something the user knows (i.e., PIN). In general, multi-factor authentication methods should be used on higher risk systems like financial transactions. Furthermore, financial institutions should be sensitive to the fact that proper implementation is key to the reliability and security of any authentication system. For example, a poorly implemented two-factor system may be less secure than a properly implemented single-factor system. The success of a particular authentication method depends on more than the technology. It also depends on appropriate policies, procedures and controls. An effective authentication method should have customer acceptance, reliable performance, scalability to accommodate growth, and interoperability with existing systems and future plans.

2. ELECTRONIC BANKING

Online banking in general is of two types: [1] Internet Banking and Electronic Banking. Internet Banking is conducted through a personal computer that can connect to a bank's website through the Internet. As an example, a customer can access a bank's website from his home using a modem and a phone line or through other telecommunications connection via an Internet Service Provider (ISP). This type of banking can also be conducted through wireless technology using the PDA or a cell phone. Electronic banking on the other hand is usually conducted using Automated Teller Machines (ATMs), telephones or debit card, but not necessarily through the Internet. Further discussions on this subject will focus on electronic banking.

E banking is defined as the automated delivery of new and traditional banking products and services directly to customers through electronic, interactive and communication channels. [3] E-banking includes the systems that enable financial institution customers, individuals or businesses, to access accounts, transact business, or obtain information on financial products and services through a public or private network, including the Internet. Customers access e-banking services using an intelligent electronic device, such as a personal computer (PC), personal digital assistant (PDA), automated teller machine (ATM), kiosk, or Touch Tone telephone.

2.1 E-banking Architecture

E-banking systems can vary significantly in their configuration depending on a number of factors. Financial institutions choose their e-banking system configuration, including outsourcing relationships, based on four factors: [8]

- the objectives for e-banking;
- the scope, scale, and complexity of equipment, system, and activities;
- the technology expertise; and
- the security and internal control requirements.

Electronic banking, and other types of online banking, offers advantages such as speed of banking, improved efficiency, and convenience as well as less paper work. But since the Internet is a public network, it presents some privacy and security problems. In general online banking poses significant risks both to the financial institution as well as the customers. Electronic banking depends on a networked environment. Network access can be performed through a combination of devices such as personal computers, telephones, interactive television equipment, and card devices with embedded computer chips. The connections are completed primarily through telephone lines, cable systems, and in some instances wireless technology. These systems whether informational or transactional, facilitate interaction between the bank and the customer, often with the support of third party service provider. It is not all networks that carry the same level of risks, and not all networks are equally vulnerable.

Banks should be careful of insider attack, which is potentially the most damaging because the bank's personnel, which can include consultants as well as employees, may have authorized access to critical computer resources. Combined with detailed knowledge relating to the bank's practices and procedures, an internal attacker could access value transfer systems directly or exploit trusted relationship among networked computers to gain a level of access that allows him to bypass established security controls. After that, the attacker could potentially transfer money or other assets inappropriately. For this reason, the first thing a bank should do is to review and evaluate the security of its internal network.

E-banking involves three main security problems: [2]

1. **Spoofing**: this involves a method of assuring customers that they are doing business with the right bank.
2. **Electronic Eavesdropping**: this involves a method of assuring customers that their account number information is not accessible to online eavesdroppers.
3. **Data Alteration**: this involves a method of assuring customers that eavesdroppers cannot alter their personal information. These can be achieved through: authentication – to guard against spoofing; privacy – to guard against eavesdropping; data integrity – to prevent data alteration; and non-repudiation – to guard against denial of a previous act.

3. AUTHORIZATION AND AUTHENTICATION

In security engineering, and computer security in general, authorization is a part of operating system that protects computer resources by allowing those resources to be used by resource consumers that have been granted authority to use them. Resources include data, files, computer programs, computer devices and functionalities provided by computer applications.

The authorization process is used to decide if person, program or device K is allowed to have access to data, functionality or service J. Most modern, multi-user operating systems include an authorization process. This makes use of the authentication process to identify consumers. When a consumer tries to use a resource, the authorization process checks that the consumer has been granted permission to use that resource. Permissions are generally defined by the computer's system administrator in some types of "security policy application", such as an access control list or a capability, on the basis of the " principle of least privilege": consumers should only be granted permissions they need to accomplish just their jobs. Older and single user operating systems often had weak or non-existent authentication and authorization systems.

"Anonymous consumers" or "guests", are consumers that have not been required to authenticate. They often have very few permissions. On a distributed system, it is often desirable to grant access without requiring a unique identity. Familiar examples of authorization tokens include keys and tickets: they grant access without proving identity. There is the concept of "trusted" consumers. Consumers that have been authenticated and are indicated as trusted are allowed unrestricted access to resources. "Partially trusted" and guests are subject to authorization for their use of protected resources. The security policy applications of some operating systems, by default, grant full access to all consumers to all resources. Others do the opposite, insisting that the administrator takes deliberate action to enable a consumer to use each resource.

Even when authorization is performed by using a combination of authentication and access control lists, the problems of maintaining the security policy data is not trivial, and often represents as much administrative burden as proving the necessary user identities. It is often desirable to remove a user's authorization: to do this with security policy application requires that the data be updateable. Website authorization and membership management systems often involve the use of Java or JavaScript code which exists in the client site HTML source code. (Example: AuthPro) Drawbacks to such systems are the relative ease in bypassing or circumventing the protection by switching off JavaScript and Meta redirects in the browser, thereby gaining access to the protected web page.

Others take advantage of server-side scripting languages such as JSP, ASP or PHP to authenticate users on the server before delivering the source code to the browser. Some systems take advantage of technology in which web pages are protected using such scripting language code snippets placed in front of the HTML code in the web page source saved in the appropriate extension on the server, such as .jsp, .asp or .php. For additional security, many of the larger websites like Yahoo and Google completely obfuscate any reference to file names in the URL that appears in the address window of the browser. In public policy, authorization is a feature of trusted systems used for security or social control. In banking, an authorization is a hold placed on a customer's account when a purchase is made using a debit card or credit card.

4. ELECTRONIC AUTHENTICATION

Verifying the identities of customers and authorizing e-banking activities are integral parts of e-banking financial services. Since traditional paper-based and in-person identity authentication methods reduce the speed and efficiency of electronic transactions, financial institutions have adopted alternative authentication methods, including:

- Passwords and personal identification numbers (PIN)
- Digital certificates using public key infrastructure (PKI)
- Microchip based devices such as smart cards or other types of tokens,
- Database comparisons e.g. fraud screening applications, and
- Biometric identification devices.

The authentication methods listed above vary in the level of security and reliability they provide and in the cost and complexity of their underlying infrastructures. As such, the choice of which technique(s) to use should be commensurate with the risks in the products and services for which they control access.

The Electronic Signatures in Global and National Commerce (E-Sign) Act establishes some uniform federal rules concerning the legal status of electronic signatures and records in commercial and consumer transactions so as to provide more legal certainty and promote the growth of electronic commerce. The development of secure digital signatures continues to evolve with some financial institutions either acting as the certification authority for digital signatures or providing repository services for digital certificates.

5. DIGITAL SIGNATURE ENCRYPTION TO E BANKING

The digital signature is a computerized equivalence of the conventional handwritten signature designed to make transactions between a bank (or receiver) and her customer (or sender) faster, smoother and secure [4, 5]. Two important properties of the current "paper mail" system must be preserved: (a) messages are private, and (b) messages can be signed. A message can be signed digitally using a privately held decryption key. Anyone can verify this signature using the corresponding publicly revealed encryption key. Digital signatures cannot be forged, and a signer cannot later deny the validity of his signature. This has obvious applications in electronic funds transfer and electronic mail systems. We examine the concept of digital signature using the public key encryption. In this method, two keys are kept by the customer namely a public key and a private key. While the public key is made known to the bank, the customer keeps the private key secret. By encrypting the instruction or the message sent by a customer's computer system, the bank is able to determine the authenticity of the supposed customer and his message. The bank would thus be in position to decide whether or not to grant the request of the customer.

5.1 General Concept of Digital Signature

For user B to send a signed message M to user A in a public key cryptosystem, he first computes his signature, S for the message M using D_B (user B's private or decryption key) [6, 7]:

$$S = D_B(M). \ldots\ldots\ldots\ldots (1)$$

Deciphering an unenciphered message is reasonable by property (d) of the public key cryptosystem: each message is the ciphertext for some other message). User B then encrypts S using E_A (i.e., user A's public key), and sends the result E_A to user A. He needs not send M as well since the receiver (user A) can compute it from S.

The recipient (user A) first decrypts the ciphertext with D_A to obtain S. User A knows who is the presumed sender of the message (user B in this case); this can be given if necessary in plaintext attached to S. The receiver then extracts the actual message with the encryption procedure of the sender, E_B in this case (available on the public file).

$$M = E_B(S)\ldots\ldots\ldots\ldots\ldots (2)$$

The recipient (user A) now possesses a message – signature pair (M, S) with properties similar to those of signed paper document. The sender cannot later deny having sent this message, since no one else could have created $S = D_B (M)$, since D_B is his private key.

The receiver can convince a "judge" that

$$E_B (S) = M,$$

and thus has proof that the signed message came from the sender. The recipient clearly cannot modify M to a different version M', since this would involve creating a corresponding signature

$$S' = D_B(M') \text{ as well.}$$

Therefore user A has received a message signed by user B, which user A can prove that user B sent, but which user A cannot modify. (Nor can user A forge user B's signature for any other message). Therefore electronic signature is not only message dependent but also signer dependent.

5.2 Applications to E-Banking

Suppose a customer (sender), S wants to open an account with a bank (receiver), R. Then the customer would essentially choose two keys namely a public key and a private key such that the latter is kept secret by the customer while the former is given to the bank. Anytime the customer contacts the bank to establish a transaction, the bank would secretly choose a random number and encrypt it using the alleged customer's public key. The bank would then ask the supposed customer to send back to it the chosen random number in order to verify the customer's signature.

If the 'customer' were not the owner of the account in question, then he would not be able to decrypt the bank's message since he does not know the decryption key. In making this method of financial transaction even more secure, the following measures are considered:

(a) The bank chooses a different random number at different transactions.
(b) The bank asks its customer to send the following information along with the message:
    (i) Time of transmission
    (ii) Date of transmission
    (iii) A password
    (iv) A sequence number
    (v) A checksum of the plaintext which includes the time, date and sequence number.

This way the transaction between a customer and a bank is further strengthened.

6. CONCLUSION

This paper presented a mechanism which uses digital signature encryption to authenticate and authorize transactions between a financial institution and her customers. Most financial institutions in Nigeria are reluctant to engage in E banking transaction out of fear of fraud. E transactions facilitate and enhance the businesses of financial institutions with their customers resulting in higher profit and convenience. Digital signature encryption assures banks and other financial institutions of the security of their E transactions and allays fears associated with fraud in E banking system.

REFERENCES

[1] Federal Trade Commission, www.ftc.gov, August 2012.
[2] E Banking: Risks and Responses, Carol Sergeant Director, Banks and Buildings Societies Financial Services Authority, www.fsa.gov.uk.
[3] Saleh M. Nsouli; Challenges of the "e banking revolution"; Finance and Development, Washington; Sep. 2002; vol.39, Iss. 3, Pg 48, 4 pgs.
[4] Adewale S.O and Falaki S. O., 2001 Electronic Commerce – Promises, Threats, Trust and Payment Systems, COAN Conference Series, Vol. 12, Pp. 84 – 103.
[5] Okereke, G. E. 2003, Security Mechanisms in Electronic Funds Transfer, M.Sc. dissertation, Department of computer Science University of Nigeria, Nsukka.
[6] Rivest, R. L., Shamir, A., Adleman, L.A.," *A Method for Obtaining Digital Signature and Public Key Cryptosystem", Communications of the ACM Vol. 21, 1978, pp. 120 – 126.*
[7] Zhu, Dan, 2002 Security Control in Inter-bank Fund Transfer, *Journal of Electronic Commerce Research, VOL. 3, NO. 1, 2002*
[8] Okereke G. E. (2006). "The Need to Adopt Electronic System of Funds Transfer in Nigeria"; *Journal of Science and Technology Research; Vol. 5, No. 2, Pp. 60 - 67.*

Authors' Briefs

Engr. Dr. George E. Okereke is a Senior lecturer at the Department of Computer Science, University of Nigeria, Nsukka. He has B.Eng. in Computer Science & Engineering from Enugu State University of Science and Technology, MSc. in Computer Science and Ph.D in Electronic Engineering both from University of Nigeria, Nsukka. His research interests are security measurements, security engineering, computer forensics, web programming and computer architecture. He has published several research and technical papers and is a member to several professional bodies.

Obianuju Assumpta Ezugwu is a lecturer at Computer Science Department University of Nigeria, Nsukka. She has both BSc and MSc in Computer Science from University of Nigeria. She is a Research Fellow in the same Department. Her research interests are Software development and Cloud computing. She is a member of several professional bodies.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Students' Information-exchange and Repository System for Nigerian Tertiary Institutions

B.C. Molokwu
Department of Computer Science
University of Nigeria
Nsukka, Nigeria
E-mail: bcmolokwu@gmail.com
Phone: +2348068975051

M.N. Agu
Department of Computer Science
University of Nigeria
Nsukka, Nigeria
E-mail: monica.agu@unn.edu.ng
Phone: +2348039329480

ABSTRACT

An eclectic observation of the student communities in most Nigerian tertiary institutions apparently reveals the absence of a framework that synchronously harmonizes the academic views, ideas, interests, goals, etc of the students. Thus this may relatively constitute a negative impact on the educational advancement and development of these student communities. In this era of rapid globalization, the impact of information and communications technology (ICT) advancement in Nigeria is having numerous positive implications on students, educators, lecturers, administrators, etc. This paper proposes the design, development and implementation of a Students' Information-exchange and Repository System (SIRS) for undergraduates and postgraduates; aimed at promoting the academic and personal development of students. The system is targeted at harnessing and utilizing the immense power of ICT techniques to create a striking impact in the way students act, think, relate and communicate with each other. Also, the system avails students the opportunity to create, share, exchange information and ideas as well as user-generated contents within a virtual community over a network of computers. The system provides information to students with respect to career development and alumni-related activities. The goal of the system is tailored at establishing an efficient and effective information-sharing network and communication channel to promote students' academic growth and excellence in all ramifications using the Internet as its backbone.

Keywords- SIRS, Information-exchange System, Repository System, Student, Alumni

African Journal of Computing & ICT Reference Format:
B.C. Molokwu &M.N. Agu (2014). Students' Information-exchange and Repository System for Nigerian Tertiary Institutions.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 53-.60.

1. INTRODUCTION

According to Cambridge Business English Dictionary [2], information-exchange is the act of people, companies, and organizations passing information from one to another, especially electronically, or a system that allows them to do this. Also, the American Heritage® Dictionary of the English Language [1] defines a repository as a place where things may be put for safekeeping or a warehouse. Thus Information-exchange and Repository System is an internet-based system application which interconnects and interlocks students around the globe with the primary aim of promoting the creation and exchange of ideas, user-generated contents, information and materials – all of which are relevant to the educational growth and development of students.

Additionally, the application synchronously maintains up-to-date data storage of students and alumni profiles; thus boosting their odds of employment opportunities in the labour market. Furthermore, in looking at how the system application influences students and how students influence the system application; it is necessary to define an Information-exchange and Repository System as a communication tool. The system allows users (students) to communicate with those they would not usually have regular contact with. For example, students living outside the campus can easily share a textbook, lecture note, project ideas, hypermedia materials, multimedia materials, etc with their colleagues living within the campus and vice versa.

This information-exchange process includes – and is not limited to – hypermedia materials, multimedia materials, electronic books, information, news, ideas, pictures, technology, personal contact data, cultural practices, and anything and everything else that people can communicate to each other in face-to-face or one-on-one interactions. Communication is no longer limited. Ideas, images and resource materials are becoming more rapidly transmitted from one place to another and are changing the way people live and experience their lives. Students can become more experienced and knowledgeable about academics, researches, findings, human society and its ways if they so choose to fortify themselves with this global flow of ideas and information. Persons are no longer limited by the time/space compression – much credits to the Internet and more specifically, greater acknowledgment to information-exchange and repository networks.

Information-exchange and repository network systems play a crucial role in the true definition of globalization. Globalization includes mobility of ideas and information. So many things which were not possible many years ago are now made possible. Information-exchange and repository networks make the world much more connected and interesting; indeed they aid in the creation of a global economy. Thus Garrett [4] stated: "No matter how many different numbers are presented or how frequently one hears them, the growth of international economic activity in the past thirty years remains staggering". Information-exchange and repository network system is something that is gradually being appreciated and employed in the educational world. The technology that enables this new level of inter-connection is a vehicle; one that has the potential to open up information to more persons than has ever been possible at any point throughout human history.

2. PROBLEM STATEMENT AND OBJECTIVES

A random survey and observation of the student communities present in most tertiary institutions of Nigeria reveals the following:

(i) Absence or deficiency of a centralized framework that unites the academic views, ideas, interests, goals, etc of the students.
(ii) Absence or deficiency of a centralized framework that offers real-time, electronic educational-enhancement opportunities aimed at complementing and promoting the competence, theoretical and practical skills of the students.
(iii) Absence or deficiency of an alumni repository system and career development platform for the students.

Thus the objectives of this paper are as stated below:

(i) Provision of a centralized network platform where educational files and materials (audio, video, picture, or text format) are stored on a central computer (or a shared server); and these resources can easily be requested, reviewed, and downloaded by only authorized users.
(ii) Provision of an alumni repository network system and career development platform which is aimed at boosting the odds of students in securing jobs at the completion of their academic programmes.
(iii) Promoting information-exchange amongst students using a common language (lingua franca) understood by all. Thus this breaks down communication barrier, caused by ethnicity difference, amongst students.

3. METHODOLOGY AND DESIGN

The general-interview-guide approach was used to extract first-hand information by randomly interviewing some groups of student from six (6) tertiary institutions in Nigeria. These institutions are as stated below:

(i) Anambra State University
(ii) College of Education, Okene
(iii) Federal University of Technology, Minna
(iv) University of Abuja
(v) University of Benin
(vi) University of Nigeria

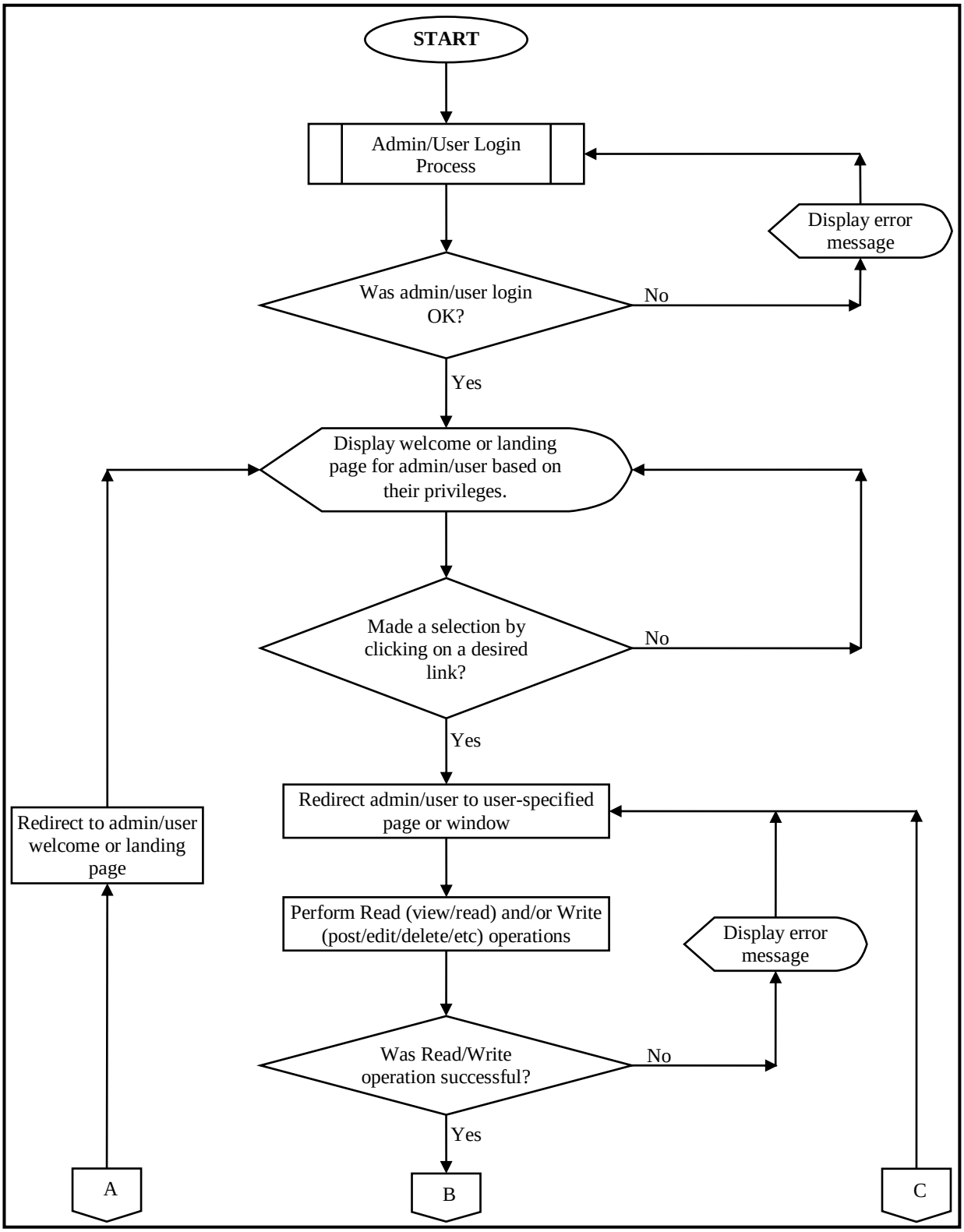
Thus our findings expose the need for the implementation of a Students' Information-exchange and Repository System (SIRS) in tertiary institutions of Nigeria for the benefit of the student communities.

Table 1 below lists the particulars which are tied to each user for identity management.

Table 1: Basic Particulars for System Users

S/N	Field Name	Data Type	Unique
1.	id	INTEGER	Yes
2.	user_id	INTEGER	Yes
3.	last_login	DATETIME	No
4.	surname	TEXT	No
5.	other_names	TEXT	No
6.	religion	TEXT	No
7.	marital_status	TEXT	No
8.	gender	TEXT	No
9.	state_of_origin	TEXT	No
10.	nationality	TEXT	No
11.	day_of_birth	INTEGER	No
12.	month_of_birth	INTEGER	No
13.	year_of_birth	INTEGER	No
14.	username	TEXT	Yes
15.	password	TEXT	No
16.	photo_link	TEXT	No
17.	study_programme	TEXT	No
18.	study_degree	TEXT	No
19.	entry_year	INTEGER	No
20.	entry_number	INTEGER	No
21.	reg_number	TEXT	Yes
22.	faculty	TEXT	No
23.	department	TEXT	No
24.	phone_number	TEXT	Yes
25.	email	TEXT	Yes
26.	hostel_address	TEXT	No
27.	home_address	TEXT	No
28.	next_of_kin_name	TEXT	No
29.	next_of_kin_phone_number	TEXT	No
30.	newsletter	BOOLEAN	No
31.	status_flag (user account validation by admin?)	BOOLEAN	No
32.	admin_check (is the user an administrator?)	BOOLEAN	No
33.	authorizer (ID of authorizing administrator)	TEXT	No
34.	eff_date (date of creation)	DATETIME	No

An overview of the general application usage for both administrators and basic users, with user-account login as a prerequisite, is as represented in the flow-chart (Figure 1) below:



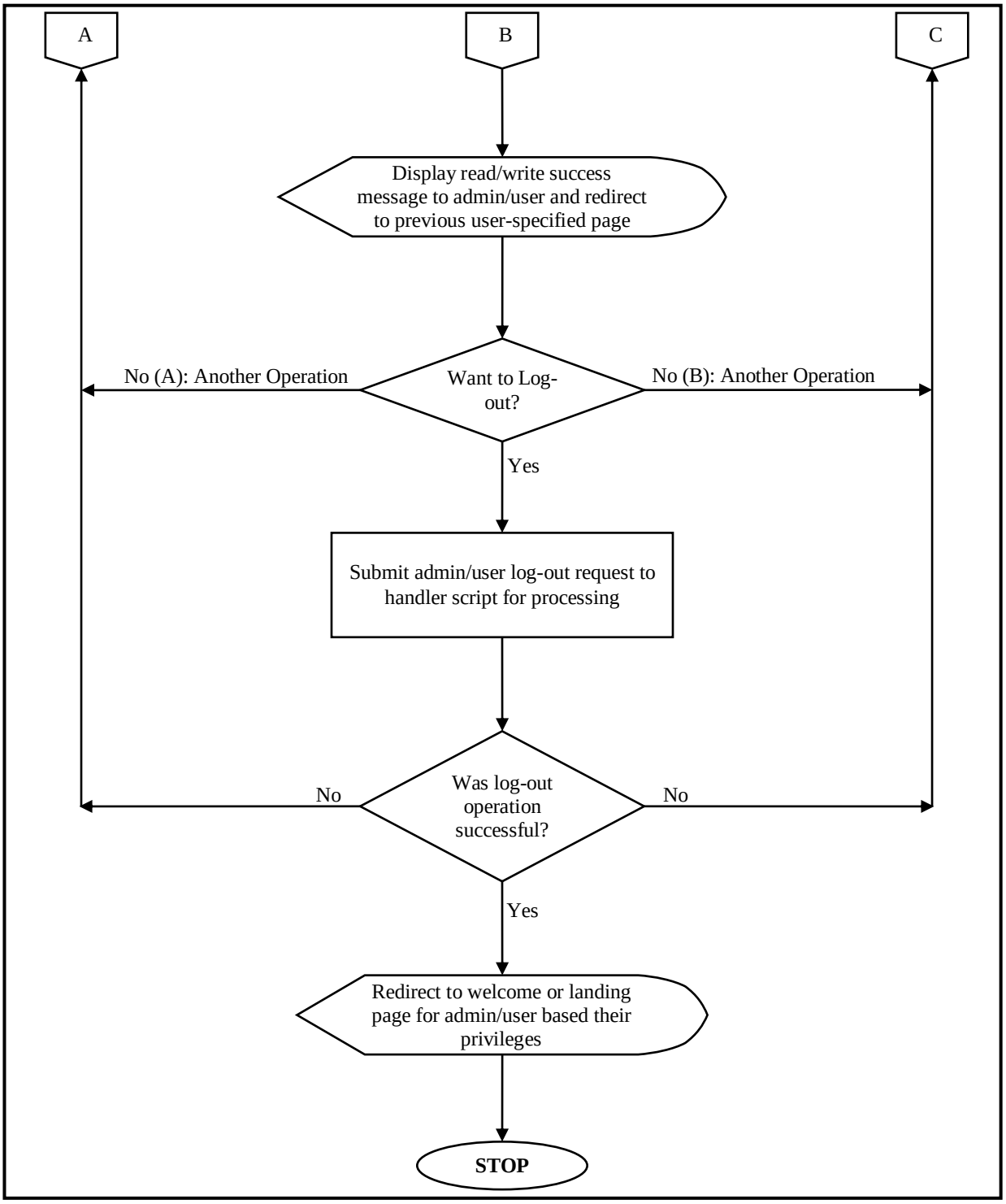


Figure 1: General Application Usage Flow-chart for Admin/User (Prerequisite = Account Login)

Additionally, another overview of the general application usage for both administrators and basic users, which does not require user-account login as a prerequisite, is as represented in the flow-chart (Figure 2) below:

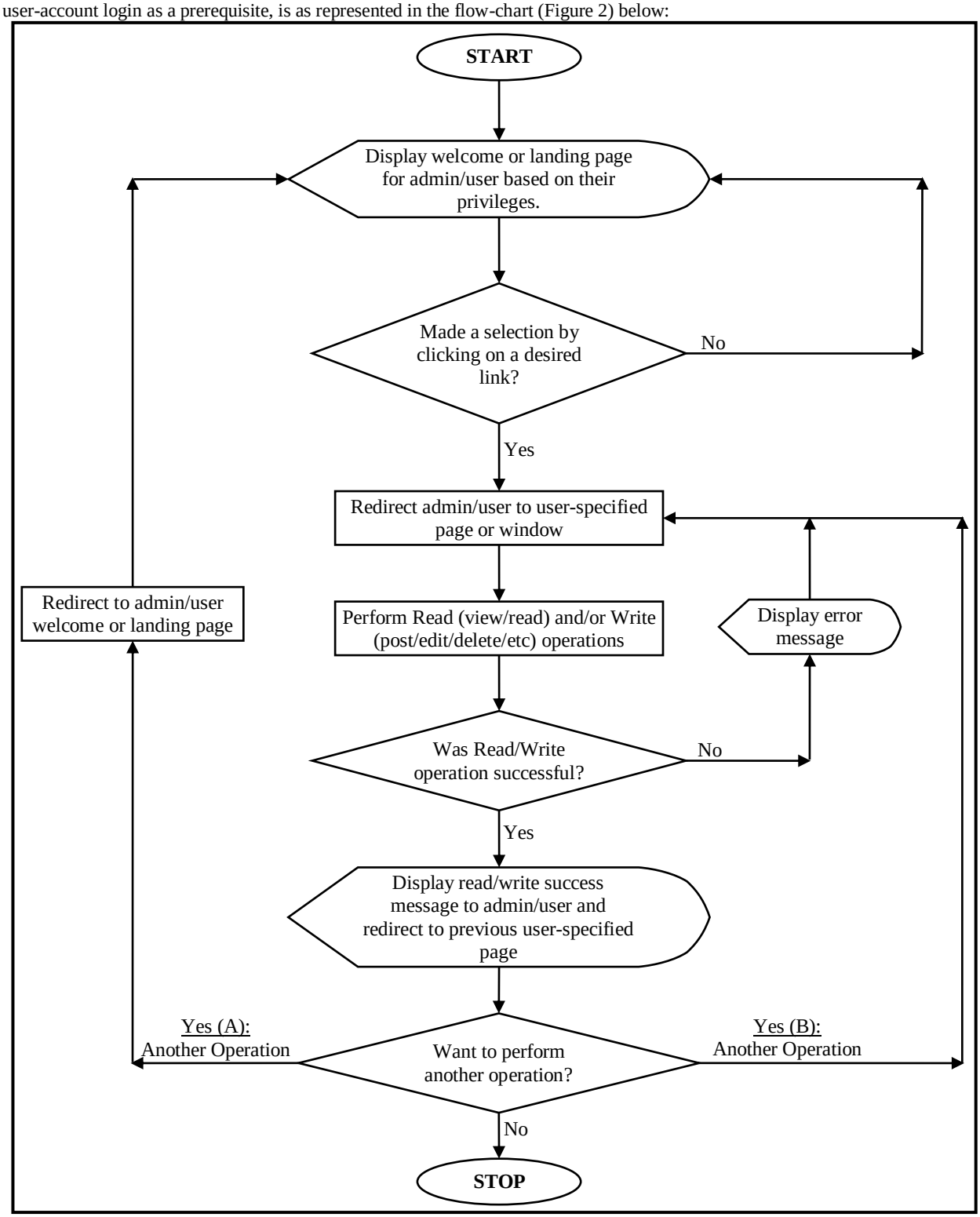


Figure 2: General Application Usage Flow-chart for Admin/User (Prerequisite = Open)

4. DEVELOPMENT AND IMPLEMENTATION

This phase of the system application development takes into consideration the following steps stated below:

4.1. Hardware and Software Interfaces

Table 2: Hardware/Software Interfaces

Minimum Hardware Requirements		
S/N	**Server-Side Specification**	**Client-Side Specification**
1.	2GHz+ CPU speed	1GHz+ CPU speed
2.	2GB+ of RAM	512MB+ of RAM
3.	10GB+ database space	512MB+ hard-drive space
4.	20GB+ hard-drive space	Internet connectivity (400kbps+ bandwidth)
5.	Web server (HTTP/FTP)	Personal Computer (PC)
6.	Database server (SQL)	
7.	Webmail server (SMTP)	
Minimum Software Requirements		
S/N	**Server-Side Specification**	**Client-Side Specification**
1.	Linux® or Windows® OS	Linux® or Windows® OS
2.	Apache™ 2.0+ server (PHP 5.0+)	JavaScript-enabled web browser
3.	MySQL™ 5.0+ server	
4.	Mercury™ 4.0+ server	

4.2 Communication Interfaces

(i) TCP/IP (Transmission Control Protocol/Internet Protocol);
(ii) HTTP (HyperText Transfer Protocol);
(iii) HTTPS (Secured HyperText Transfer Protocol);
(iv) FTP (File Transfer Protocol);
(v) SMTP (Small Mail Transfer Protocol); etc.

4.3 Software Development Models

(i) Waterfall model; and/or
(ii) Rapid Application Development (RAD) model.

Software Development and Implementation

Table 3: Software Development/Implementation

S/N	**Front-end Specification**	**Back-end Specification**
1.	HTML 4.0+	MySQL 5.0+
2.	CSS 2.0+	PHP 5.0+
3.	JQuery 1.8+	JQuery 1.8+
4.	JavaScript 1.5+	JavaScript 1.5+
5.	Adobe Photoshop 7.0+	RSS 2.0+

4.4 System Maintenance

Routine maintenance checks and application upgrade shall be the full responsibility of the web administrator. The database administrator shall be responsible for performing routine database backup/recovery and optimization operations on the database server to boost its effectiveness and efficiency.

4.5 User Characteristics

(i) **Technical Expertise:** Users ought to be computer literates and be able to use web application.
(ii) **Educational Level:** Users must, at least, be an undergraduate student; a basic understanding of English Language is very necessary.
(iii) **Experience:** Users must be conversant with his/her institution's rules and code of conduct.

4.6 Constraints

(i) Additional network-security strategies or policies have to be imposed on the system application when it is deployed to a live server. This will protect vital data and information from unauthorized application/network access (data theft).
(ii) To improve application security, it may be helpful to deploy the system application on an intranet or extranet. However, the application can also be deployed on the Internet.

4.7 Assumptions and Dependencies

(i) The system application is intended for student-related educational and career activities.
(ii) Users of the system are bona fide students of the institutions which operate on the application.
(iii) All users are computer literates.

5 FINDINGS AND DISCUSSIONS

Thus some of the goals and expected outcomes of the system application with respect to students' communities in Nigerian tertiary institutions are outlined below:

- Promote collaboration amongst students so as to upgrade their theoretical and practical skills without worrying about any physical boundaries.
- Provision of a centralized network platform where educational files and materials can be stored in a shared server, and these resources can easily be requested, reviewed, and downloaded by the users.
- To increase participation amongst students in the creation and sharing of information, knowledge, ideas, opinions, etc.
- Creation and development of "self-educating" communities: where students are able to converse with their colleagues whom share same interests and views as themselves.
- Break-down of language barriers in a bid to harmonize the different socio-cultural groups in Nigeria; thus this makes the students feel much more connected and united.

- Promotion of globalization because the students now form newer relationships that can transcend geographic boundaries and barriers.
- Provision of an alumni repository network system for the management of students' data or information.
- Provision of a career development platform aimed at boosting the odds of students in securing jobs after the completion of their academic programmes.

6 CONCLUSION AND FUTURE WORKS

This paper has been able to present the design, development and implementation of a Students' Information-exchange and Repository System (SIRS) for adoption in Nigerian tertiary institutions. The system application harnesses the power of information and communication technology innovations in creating and developing a platform aimed at boosting the knowledgebase and productivity of students (users). Future work will be concentrated on broadening the scope of the system to serve as a standard platform for interconnecting all students in different tertiary institutions of Nigeria. In that regard, it aims at creating a much wider network of student communities.

REFERENCES

[1] American Heritage® Dictionary of the English Language, 2009. Repository. Fourth edition, Houghton Mifflin Company Publishers.
[2] Cambridge Business English Dictionary, 2014. Information Exchange. Cambridge University Press.
[3] C. Burrus, 2014. Open Educational Resources (OER) and Connexions. Rice University, Houston, Texas, USA.
[4] Encarta Encyclopedia, 2009. Science and Technology. Microsoft Corporation.
[5] G. Garrett, 2000. The Causes of Globalization. Yale University, New Haven, Connecticut, USA: 1-10.
[6] H. Deitel, P. Deitel, T. Nieto, 2001. Internet & World Wide Web: How to Program. Second Edition, Deitel & Associates Incorporated.
[7] J. Greenspan, B. Bulger, 2001. MySQL/PHP Database Applications. IDG Books Worldwide Incorporated, California, USA.
[8] J. Owyang, 2009. The Many Challenges of Social Network Sites. Available: http://www.web-strategist.com/blog/2008/02/11/the-many-challenges-of-social-networks/.
[9] L. Smolin, N. Burbules, K. Lawless, 2007. Information and Communication Technologies: Considerations of Current Practice for Teachers and Teacher Educators. Blackwell Publishers, New York, USA.
[10] L. Welling, L. Thomson, 2001. PHP and MySQL Web Development. Sams Publishing, Indianaplolis, Indiana, USA.
[11] N. Burbules, 2006. Self Educating Communities: Collaboration and Learning through the Internet.
[12] T. Mitrano, 2006. A Wider World - Youth, Privacy, and Social Networking Technologies. Educause Review, Nov/Dec, 16-28.
[13] Z. Bekerman, N. Burbules, D. Silberman-Keller, 2006. Learning in Places: The Informal Education Reader. Peter Lang Publishing, New York, USA.

Author's Biography

Bonaventure C. Molokwu is a certified programmer and web-application developer as well as a member of a couple of international Information and Communications Technology societies. He started his career in programming and still finds it most interesting. He has over six years of experience in software-application development and programming: ranging from small-scale to enterprise applications. At the moment, he is a postgraduate of Computer Science at University of Nigeria, Nsukka. He holds a Bachelor of Science degree in Computer Science from University of Benin, Benin City, Nigeria. His areas of research include: Software Engineering and Security, Intelligent Systems and Networking. He can be reached via phone on +2348068975051 and through E-mail bcmolokwu@gmail.com

Dr(Mrs) Monica N. Agu is a Senior Lecturer at the Department of Computer Science, University of Nigeria, Nsukka. She obtained a BSC Computer Science at the University of Ibadan in 1978, Master of Science Degree in Computer Science at the University of Nigeria, Nsukka in 1995 and PhD Degree in Computer Science from Ebonyi State University, Abakaliki in 2009. Her research has focused on using Information and Communication Technology to improve the lives of people and Modelling performance of Computer Systems. She can be reached by phone on +2348039329480 and through E-mail monica.agu@unn.edu.ng

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Algorithm Development for Scheduling Unemployment Reduction in Nigeria

M. Olagunju
Department of Computer Science
Institute of Information and Communication Technology
Kwara State Polytechnic
Ilorin, Kwara State, Nigeria.
olamukaila.yahoo.com

A. Ajao
Department of Computer Engineering
Federal Polytechnic
Offa, Kwara State Nigeria.
ajaodoc@gmail.com

ABSTRACT

The major problem facing most of the countries of the world is unemployment of youth after the graduations from the institutions. This problem normally occur in the under developed and developing countries of the world in which Nigeria is included. Most of these undeveloped and developed nations are not planned ahead for provision of employment for their youth after their graduation. This paper used Nigeria census of 2006 in some selected states spreading from six geographical zones, that is the Southwest, South South, SouthEast,NorthCentral,NorthWest and NorthEast in which some state were chosen from each geographical zone and based on their population census of 2006. The analysis was carried out to determine the unemployment reduction plan for each state. The main objective of this study is to develop the algorithm that can schedule the unemployment reduction in Nigeria. The algorithm was developed with modification of Mersenne Twister. The algorithm was used in the program that was developed with MatLab to predict rate of unemployment reduction scheduling. This will definitely assist the government and other stake holders to plan ahead for unemployment reduction process.

Keywords: Algorithm, Unemployment, Scheduling, Development etc.

African Journal of Computing & ICT Reference Format:
M. Olagunju & A. Ajao (2014). Algorithm Development for Scheduling Unemployment Reduction in Nigeria.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 61-68. .

1. INTRODUCTION

The economy rate of a country is said to be extremely poor when the rate of unemployment is greater than the rate of employment in the country. Unemployment is simply defined as the total number of able men and women of working age seeking for a paid work [1]. Nigeria is the most populous country in Africa and the eight in the world with a population of over 140 million people by 2006 census. When the rate of people seeking for a paid work is much in a country which leads to unemployment the country will automatically be under the pressure of poverty. Unemployment and poverty are so intertwine that one can easily confuse one for the other. Although, it is possible for one to be employed and still poor, this is likely to be a case of underemployment. Thus, by unemployment, it includes those underemployed. Unemployment and underemployment reflect the failure to make use of an important factor of production, labor, for fostering economic growth in Nigeria. Low returns to labour as well as high unemployment indicates poverty. Poverty makes it difficult to make investments in education and health that would increase a person‘s productivity.

The inadequate employment situation of youth has a number of socio-economic, political and moral consequences. This has resulted in poverty in Nigeria which is chronic and rising. The share of the total population living below the $1 a day threshold of 46 per cent is higher today than in the 1980s and 1990s. Unemployment is the greatest challenge to underdeveloped and developing countries. the phenomenon of graduate unemployment as it is being experienced in the developing countries constitute a peculiar problems to labor market and the general economy of these countries [5]. From the content analysis perceptions of job seekers on the issue of graduate unemployment in a study conducted by [7], the following factors were identified as the major causes of unemployment in Nigeria: The long period of initial unemployment among university graduates in Nigeria, faulty manpower planning and expansion of educational facilities that have unduly raised the expectations of Nigerian youths, the economic recession, continued proportionality of expatriates in employment, the institution of NYSC, the collective bargaining process, graduate attitude to some type of jobs attitude to jobs in other location as well as search behavior of employers and job seekers, use of capital intensive technology , wide rural urban migration , formal informal sectors differentials.

All these and many other factors contribute the causes of graduate unemployment in Nigeria. Unemployment in Nigeria is one of the most critical problems the country is facing. The years of corruption, civil war, military rule, and mismanagement have hindered economic growth of the country. Nigeria is endowed with diverse and infinite resources, both human and material. However, years of negligence and adverse policies have led to the under-utilization of these resources. These resources have not been effectively utilized in order to yield maximum economic benefits. This is one of the primary causes of unemployment and poverty in Nigeria. Another major causes of unemployment is Nigeria is our Educational Sector. Educational sector is one major cause in the sense that there are some certain things we need to put in other, that is to say to correct our educational curriculum, other west Africa countries writing the WEST AFRICA EXAMINATION COUNCIL (WAEC) are offering some hand work activities, Activities like bricklaying, tailoring, textile and design etc. Unemployment in Nigeria is a major problem both economically and socially. Unemployment in Nigeria has resulted in more and more people who do not have purchasing power. Less consumption has led to lower production and economic growth has been hampered. Unemployment also has social consequences as it increases the rate of crime.

1.1 Solutions to the problem of Unemployment

Economic growth is not the only solution to curb unemployment in Nigeria, as the official statistics illustrate that previously unemployment did not always decline with the economic growth. Other solutions such as the provision of right skills to the people to help them tackle the problems and lead a more prosperous life should also be given importance. The following suggestions may play a cogent role in eliminating unemployment in Nigeria.

These include:-

- Sports schools, evening clubs that teach kids to play football, swim etc
- Computer training schools and clubs that specializes in teaching programming software such as Java, Oracle, ASP, Cold fusion, JSP, digital photography, and video editing, etc.
- Language schools teaching foreign languages like French, Spanish, and Chinese.
- Setting up of provisional work agencies, which provides temporary staff to small companies
- People with good web and programming skills can think of starting the following projects in order to deal with the problems of unemployment - program unique JAVA based applications to be used in 3G phones, a project to structure a complete phone directory of all Nigerian phone numbers, online map project of major Nigerian cities, and project, which can provide sufficient information about everything in Nigeria.
- Other projects such as building solar powered water pump for use in rural areas, solar powered streetlights, hybrid powered generators using solar and battery power, can also generate employment in the Nigerian economy.

Computer simulation is the discipline of designing a model of an actual or theoretical physical system, executing the model on a digital computer, and analyzing the execution output. For the simulation to be more effective it must be tailor towards and effective algorithm schedule. Algorithm Schedule is the procedure of accomplishing a giving task, while scheduling is the process of assigning and appropriate number of workers to the job during each day. To solve the problem of unemployment in Nigeria, this study focuses on using computer simulation algorithm scheduling for effective reducing of unemployment.

2. REVIEW OF PREVIOUS WORKS

[7], and [6] opined that unemployment can be describe as the state of worklessness experienced by persons who are members of the labour force who perceived themselves and are perceived by others as capable of work. Unemployed people can be categorized into those who have never worked after graduation from the university and those who and those who have lost their jobs thereby seeking re-entry into labour market. However, most of the previous study on unemployment of youths especially of graduate unemployment in developing countries [2] has tended to ignore the special case of the university graduates that are first time job seeker. This study makes an attempt to stimulate and algorithm for unemployment scheduling. [4] the meaning of work to paid employment is the result of the development of capitalist productive relations. However, according to [7] the concept of work has partly shifted from productive effort itself to the predominant social relationship. For instance, it is only in the sense of social relationship that a woman running a house and bringing up children can be said not to be working [1].

3. MATERIALS AND METHOD

In this paper, the algorithm to show case the unemployment reduction is giving below with algorithm representation which is modified mersenne twister. The material use for this paper is based on Nigeria population census of 2006, in which some states where selected from the six geographical zones

3.1 Modified Mersenne Twister Algorithm for Unemployment

1. Start
2. Enter the current year, I
3. Enter the predicted year, n
4. For K= I to n
5. Enter the population of current year, g
6. Enter the rate of unemployment, f
7. Let R(rate)=f/100
8. Let P=(100-r)/100
9. Let Z=p*g
10. Let J=g-z
11. Display k, R, P, Z, J
12. Next K
13. Stop

In the algorithm above, K is the counter
R is the rate of unemployment
P is the rate of employment
Z is the population of employment
J is the population of unemployment

4. DISCUSSION OF RESULT

This section focuses on discussion of result generated with MatLab version 6.0. and the results are discussed in the table below:

Table 1: This displays the analysis of number of employment and unemployment rate in Abia state of Nigeria.
ABIA STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	2836964	421116	2415848	2415848
2	2015	5678133	3266496	2411637	5257017
3	2016	8519302	6111876	2407425	8098185
4	2017	11360470	8957256	2403214	10939354
5	2018	14201639	11802636	2399003	13780523
6	2019	17042808	14648016	2394792	16621692
7	2020	19883977	17493396	2390581	19462861

Table 2: This displays the analysis of number of employment and unemployment rate in Anambra state of Nigeria.
ANAMBRA STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	4172315	275737	3896578	3896578
2	2015	8347386	4453565	3893821	8071649
3	2016	12522456	8631393	3891064	12246720
4	2017	16697527	12809221	3888306	16421790
5	2018	20872598	16987049	3885549	20596861
6	2019	25047668	21164877	3882792	24771932
7	2020	29222739	25342705	3880034	28947002

Table 3: This displays the analysis of number of employment and unemployment rate in Cross River State of Nigeria.
CROSS RIVER STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	2888419	228546	2659873	2659873
2	2015	5779121	3121534	2657587	5550575
3	2016	8669824	6014522	2655302	8441278
4	2017	11560527	8907510	2653016	11331980
5	2018	14451229	11800498	2650731	14222683
6	2019	17341932	14693486	2648446	17113386
7	2020	20232634	17586474	2646160	20004088

Table 4: This displays the analysis of number of employment and unemployment rate in Ekiti state of Nigeria.
EKITI STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	2390568	419817	1970751	1970751
2	2015	4785327	2818774	1966552	4365509
3	2016	7180086	5217731	1962354	6760268
4	2017	9574844	7616688	1958156	9155027
5	2018	11969603	10015645	1953958	11549786
6	2019	14364362	12414602	1949760	13944545
7	2020	16759121	14813559	1945561	16339303

Table 5: This displays the analysis of number of employment and unemployment rate in Kogi state of Nigeria.
KOGI STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	3300866	659495	2641372	2641372
2	2015	6608314	3973538	2634777	5948820
3	2016	9915762	7287581	2628182	9256268
4	2017	13223210	10601624	2621587	12563716
5	2018	16530658	13915667	2614992	15871164
6	2019	19838107	17229710	2608397	19178612
7	2020	23145555	20543753	2601802	22486060

Table 6: This displays the analysis of number of employment and unemployment rate in Oyo state of Nigeria.
OYO STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	5573083	390663	5182421	5182421
2	2015	11150071	5971557	5178514	10759408
3	2016	16727058	11552451	5174608	16336396
4	2017	22304046	17133345	5170701	21913383
5	2018	27881033	22714239	5166794	27490370
6	2019	33458020	28295133	5162888	33067358
7	2020	39035008	33876027	5158981	38644345

Table7: This displays the analysis of number of employment and unemployment rate in Osun state of Nigeria.
OSUN STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	3416276	34170	3382106	3382106
2	2015	6832893	3451129	3381764	6798723
3	2016	10249510	6868088	3381423	10215341
4	2017	13666128	10285047	3381081	13631958
5	2018	17082745	13702006	3380739	17048575
6	2019	20499362	17118965	3380398	20465193
7	2020	2391597	20535924	3380056	23881810

Table 8: This displays the analysis of number of employment and unemployment rate in Ogun state of Nigeria.
OGUN STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	3744241	345105	3399136	3399136
2	2015	7491930	4096245	3395685	7146825
3	2016	11239619	7847385	3392234	10894514
4	2017	14987308	11598525	3388783	14642203
5	2018	18734997	15349665	3385332	18389892
6	2019	22482686	19100805	3381881	22137581
7	2020	26230375	22851945	3378430	25885270

Table 9: This displays the analysis of number of employment and unemployment rate in Sokoto state of Nigeria.
SOKOTO STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	3699640	151810	3547831	3547831
2	2015	7400798	3854486	3546313	7248989
3	2016	11101956	7557162	3544795	10950147
4	2017	14803114	11259838	3543276	14651304
5	2018	18504272	14962514	3541758	18352462
6	2019	22205430	18665190	3540240	22053620
7	2020	25906588	22367866	3538722	25754778

Table10: This displays the analysis of number of employment and unemployment rate in Kano state of Nigeria.
KANO STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	9377236	1203365	8173871	8173871
2	2015	18766490	10604653	8161838	17563126
3	2016	28155745	20005941	8149804	26952380
4	2017	37544999	29407229	8137770	36341634
5	2018	46934254	38808517	8125737	45730889
6	2019	56323508	48209805	8113703	55120143
7	2020	65712762	57611093	8101669	64509397

Table 11: This displays the analysis of number of employment and unemployment rate in Nasarawa state of Nigeria.
NASARAWA STATE

DEGREE OF PRIDICTION	YEAR	POPULATION	NO. OF EMPLOYMENT	NO. OF UNEMPLOYMENT	ANALYSIS
1	2014	1868779	29910	1838869	1838869
2	2015	3737857	1899287	1838570	3707947
3	2016	5606935	3768664	1838271	5577025
4	2017	7476013	5638041	1837972	7446103
5	2018	9345090	7507418	1837672	9315180
6	2019	11214168	9376795	1837373	11184258
7	2020	13083246	11246172	1837074	13053336

5. CONCLUSION AND RECOMMENDATION

We find out that with proper planning by the government and all the stake holders in Nigeria the unemployment rate especially youth with skilled acquisition can be reduce if properly monitored with the design algorithm which follows the modified mersenne twister algorithm.

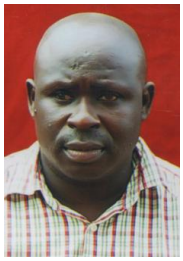
We recommend the following:

1. Proper planning must be carried out to reduce unemployment rate in Nigeria.
2. Expect must be involved in planning for the under graduate youth in order to avoid unemployment after their graduation.
3. Industrialization must be expanded.
4. Entrepreneurship skills must be encourages right from the secondary school to the higher institution in order to enhance their working capability.
5. Over dependent on oil for survive must be discourage and mineral resources in each state of Nigeria must be tap to provides employment for the graduate youth

REFERENCES

[1] Beer , W And Herves M.E.A (1966). ' Employment And Industrialization In Developing Countries', Quarterly journal of Economics, Vol.30 No.1
[2] Bhalla A.S (1973) ' A Disaggregative Approach To Employment In Less Developed Countries ' The Journal Of Development Studies
[3] Berg, E.J. (1969) 'wages policy and employment in less developed countries ', in R. Robinson, Ibid.
[4] Diejomoah V.P And Orimolade W.A.T (1971). ' Unemployment In Nigeria: An Economic Analysis Of Scope , Trends And policy Issues ' Nigerian Journal Of Economic And Social Studies Vol.13 no.2 Pp 127 – 160.
[5] Berg, E.J. (1969) 'Urban Real wage And the Nigerian Trade Union Movement (1939 – 1960) a comment' Economic Development and Cultural Change,
[6] Fajana, S. (1987) economic Recession, Collective Bargaining And labour Market Segmentation In Nigeria, Nigerian Management Review CMD, Lagos, Vol.2, No 1, P9 - 16
[7] Fajana, S (2000) Functioning Of the Nigerian Labour Market, Labofin and Company, Lagos, Nigeria

Authors' Biographies



Olagunju Mukaila is a senior lecturer and head, Computer Science Department, Kwara State Polytechnic, Ilorin, Ilorin Nigeria. Olagunju obtained his first degree in Computer Science and Mathematics (Combined Honour), from Federal University of Technology Minna, Nigeria. He obtained his second degree in Mathematics with Computer Science option from University of Ilorin, Ilorin Nigeria. He also obtained his PHD in Computer Science from University of Ilorin, Ilorin Nigeria. Olagunju is a Computer and Mathematics expert, his area of specialization includes, Algorithm development and fluid flow pattern recognition. He has published Articles both in the National and International Journals. One of his popular articles is determination of time of death predictive model. Olagunju is a member of Computer Professional Council of Nigeria (CPN) and Nigeria Computer Society (NCS)

Abdulazeez O. Ajao received the B.Sc degree in Computer Engineering from Obafemi Awolowo University, Ile-Ife, Nigeria, in 1995 and M. Eng. degree in Electrical Engineering from University of Ilorin, in 2006. Currently, He is a Principal Lecturer at the Federal Polytechnic Offa. His main research interests are in the areas of image processing, Electronics and various industrial applications.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Conceptual Model of Energy Conservation Management System in Renewable Energy Installations

A.K. Shittu (MIEEE) & A. Fasipe
School of Computing
University Utara
Malaysia

N.A. Adedokun-Shittu (MIEEE)
School of Business
University Utara
Malaysia

S.A. Opatokun
Department of Environment and Geography
Macquarie University
Australia.

Corresponding author: abdjaleel@uum.edu.my

ABSTRACT

Energy issues, more than ever have become a perennial concern in the recent times because it is the bedrock of industrial development which leads to national prosperity. The use of fossil fuel in energy production has increased the level of pollution. The challenges of energy production include high cost of energy, global warming and the resultant climate change. The future challenges of energy production, sustainability and energy security have increased the need for an alternative source of energy. The focus has been on the renewable energy sector to provide the energy of the future. The purpose of this paper is to develop a conceptual model that will enable the synchronization of energy sources and ultimately conserve energy consumption without reducing the quality of life. In the light of this, the study will help develop a conceptual model for energy conservation management system that will integrate the renewable energy installations for effective usage, and hence energy conservation. The management system includes the grouping of the household appliances and how they interact with power supply to conserve energy consumption through the distribution board. To achieve the energy conservation management system, it will compose of 10 stages. The system will be called energy conservation management system (ECMS). The first step will be to understand the needs of the users, requirement analysis, renewable technologies, materials, and energy integration and ends up with a comprehensive design plan for the user. The design phases of ECMS are in such a way that there are checkpoints. This will help the installer to check for the synchronization and conservation of energy issues of the renewable installation. The use of the proposed model to design the synchronization and conservation will reduce energy consumption and seamlessly integrate the energy sources. It is possible to improve on the model by carefully reviewing each part of the components.

Keywords: Conceptual Model; Energy Conservation; Renewable Energy; Energy Management System; Energy Optimization, & Green Technology.

African Journal of Computing & ICT Reference Format:
A.K. Shittu, A. Fasipe, N.A. Adedokun-Shittu & S.A. Opatokun (2014). Conceptual Model of Energy Conservation Management System in Renewable Energy Installations. Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 69-76.

1. INTRODUCTION

Energy is the bedrock of industrial development which leads to national prosperity. The world has used various forms of energy in developing and maintaining industries in the past [1]. Currently, fossil fuel is used in energy production. In the recent time, there has been a need to venture into other forms of energy production for industrial consumption for the future [2, 1, and 4]. The focus has been on the renewable energy sector to provide the energy of the future [12]. The renewable energy is sustainable and affordable when amortized over a period of time.

Renewable energy has the capability to reduce the greenhouse gases that harm the environment from the use of fossil fuels. This is evident when the United Nations Secretary-General, Ban Ki-moon, said that renewable energy has the ability to lift the poorest nations to new levels of prosperity [3]. Renewable energy is the type of energy that comes from a source that continually replenishes. The sources that continually replenish are sun, water, and wind. These sources occur naturally in nature [4].

The renewable energy technologies, directly or indirectly, consists of the following – solar energy, wind power, hydropower, and geothermal energy just to mention a few [5, 6]. The renewable energy is believed to be a clean energy which must not be wasted but be conserved [7]. The challenges are the ways to transit from fossil based energy to renewable energy without the sudden break in the energy supply and the conservation of energy without reducing the quality of life. The most important concern to the user is the good life before considering the issues of environmental pollution and eventual global warming. In order to overcome these challenges of this transition phase, the sources of energy should be synchronised for the benefit of the end user and eventual energy conservation.

These two challenges are more prominent when Nigeria as a country is being considered. The Nigerian situation is peculiar because of the epileptic power supply over the year. Not only Nigeria, but any country that is experiencing epileptic power supply will fall under the same category of these challenges. The challenge has been the ability to manually integrate the energy sources and conserve energy use. There is lack of energy management system to synchronise the energy production in renewable energy installations and the existing sources of energy with consumption of the consumers with the ultimate aim of conservation. The introduction of renewable energy in the current energy market has given a new dimension to the future of energy. However, the renewable energy did not exist in a vacuum. Other sources of energy from fossil fuel and nuclear energy are in existence [8]. The users of energy should be able to remotely control their consumption that will result into conservation of energy. Due to the lack of conservation, there is the problem of wastage in consumption [9]. The manual system of switches is not proofing to be effective as forgetfulness and lack of conservative culture ruins the effort.

The study aims at developing a conceptual model of energy management system that will solve the problem and increase efficiency and conserve the energy. The energy conservation management system is designed to help integrate the renewable energy installations for effective usage. The under-listed objectives will be carried out to achieve this purpose:

i. To understand the current problems encountered in renewable energy installation
ii. To construct the conceptual model for energy conservation management system.
iii. To develop a conceptual model for energy conservation management system.

The clamour for a shift into the world of renewable energy has been on the increase lately. Funds are invested into many researches in order to develop this sector. One of the major issues is the conservation of energy that is produced from the renewable energy [2, 4]. The development of a management system that will help conserve the energy during usage will go a long way in making the renewable energy sector a worthwhile investment over time.

2. ENERGY PRODUCTION

Energy is recently and increasingly becoming a scarce commodity and valuable resource irrespective of the form [9, 11]. The world must ensure reliable, affordable and clean energy for the sustainability [8]. Over the past few decades, there has been a growing dependence on fossil fuels for the generation of electricity. The fossil fuels which are in use though in different proportions are coal, oil and natural gas. Fossil fuel consumption is the major contributor to the increasing concentration of carbon dioxide (CO_2) in the atmosphere, a key cause of global warming [15]. Also, the fossil fuels release greenhouse gases and other pollutants into the atmosphere [16]. The release of the greenhouse gases has impacts on the earth and mankind in general [5].

The economic development has caused a rapid depletion of these limited natural resources [11]. The shortage will affect the activities of all walks of life and impede economic development [32]. There is a possibility of global degradation of people's living standard if the is not known alternatives when this happens [16]. The two major issues in the future of energy are energy security and climate change [36]. It is believed that the future contribution of renewable energy sources will be vital in having energy security in the future [22].

Renewable energy exists perpetually and in abundant quantity in the environment [10]. It is ready to be harnessed, inexhaustible, and importantly, it is a clean alternative to fossil fuels [23, 24, 12]. Renewable energy technologies turn these fuels into usable forms of energy—most often electricity, also heat, chemicals, or mechanical power [8, 25]. There is strong optimism that renewable energy investment is a pathway to the future's energy. Despite the high cost of initial installation, there is optimism that renewable energy will pay off economically in the long term [12, 7].

2.1 The Current Trend in Renewable Energy

Scholars have noted the exiting prospects of renewable energy, as well as the potential challenges that come with utilizing the renewable fuels [8, 13, 14]. Countries of the world are setting new regulations to implement the renewable energy policy [28]. Example of such is the Renewable Energy Directive of 209/28/EC which established a European framework for the promotion of renewable energy [35, 17]. Reducing fossil fuel consumption may slow the rate of global warming [15]. The development and use of renewable energy is expected to increase as fossil fuel supplies decline [18].

2.2 The Current Situation in the Country of Nigeria

Nigeria, a country covering a total 356,667sq miles (923,768 sq kn) in the West of African continent has 351,649 sq miles (910, 771 sq kn or 98.6% of total area) to be land. The country has approximately 160 million people. Nigeria is divided into six Geo-Political Zones and has a total of 36 states excluding the Federal Capital Territory (FCT) [19].

The energy need in Nigeria is on the increase as the nation is set on the path industrialized countries by the year 2020. However, power outage in Nigeria is a common occurrence. To salvage the current situation, the use of petrol/diesel generators has increased. Different capacities of generators are in use all over the country [20]. Some of the issues causing the low power production in Nigeria are low water levels, unavailability of gas to power the turbines, and on the overall, lack of proper maintenance of the power plants installed by the Power Holding Company nationwide [21].

The current power production ranges between 2,600MW and 3,600MW. The power production is grossly inadequate for any meaningful development in the country of approximately 160-170 million people [6]. The attitude of the workers to the situation is appalling. The workers are more confused more than ever as they plan to privatise the holding [21]. Nigeria has enough natural resources that will support its power generation. However, the natural resources have not translated to energy generation for the country. There is shortage of electrical power [33]. Only 40% of the estimated population of 160 million have access to the electricity supply. The availability is worse in the rural area where the 70% of the population resides as only 15% have access to electricity [19].

3. TRASITION FROM FOSSIL FUEL TO RENEWABLE ENERGY

In transiting from reliance on fossil fuel energy resources to renewable energy technologies, there are various combinations of energy sources at this phase. The sources should be in a way that is synchronised to allow for maximum benefit and conservation of energy. The synchronization of the electricity grid and the renewable energy installations must help to improve storage capacity [8, 12, 1]. There should be a better system controls that will help improve the efficiency use of the energy sources with the existing infrastructure without necessarily having to invest so much. If this can be achieved, this will prevent transmission losses and optimal use [12]. The field of artificial intelligence has helped to serve as a source for improving the human life. The application of artificial intelligence will remove the human agent and improve maximal synchronization of the energy sources [29].

3.1 A Need for Conservation in Energy

The population of the world is expected to grow up to 10 billion by 2050. The increase in population will lead to increase energy demand. This will result to increase of fuels and electricity prices, which is already happening [16]. There is challenge of space and initial investment in the renewable energy. Even if the potentials of renewable energy is developed, there is every tendency of not been sufficient. This means that the production of energy through the renewable energy means is definite [7]. It is important for everybody to work together in conserving energy [22]. The large-scale production of renewable energy presents some challenges. The land required for collection and production of renewable energy affects the agriculture, forestry, and urbanization [26].

It is paramount to make energy usage efficient. Technologies to reduce the consumption are by increasing energy efficiency, reducing power usage, and making use of the renewable energy [27]. For example it is estimated that the amount of energy loss through poorly insulated windows where air conditions are used in Alaska in the United states is 1.1 x 1012 kWh(3.8 quads) each year. This energy is capable of powering all the oil pumps in Alaska per year. This means that a substantial amount of energy is lost to improper action of conservation of effective use [28].

4. RELATED WORK

Various works in literatures exist in the area of energy management modelling. The structured analysis and design methodology was started in the mid-1970s. The emergence came to complement structured programming. The data flow modelling methodology has become one of the dominant methodologies that are utilized for business-oriented information systems development. The real world scenario will be described by the flow of data through the information system. The data will be transformed into information as the data move from input to storage to output. However, the structured analysis and design methodology gives a kind of dimension that will allow it to be formalized, measurable, and reputable. In addition, the computer-aided/software engineering (CASE) which was introduces in the early 1980s will give a new dimension. The dimension will allow for the enhancement of automated support. It is worthy of note that the CASE acronym has evolved to (IDE or SDE) integrated/software development environment.

5. SYSTEM DESIGN

The system will use the problem solving approach, this is by asking and answering the question “what does the system have to do?” This is referred to as the functional perspective approach. The tools of CASE/IDE will provide assistance in drawing the methodology notations, validation and verifying the methodology models. The problem solving will approached the problem from an object perspective. Therefore, object-oriented systems analysis and design approach that will lead to programming task will be employed. The eight characteristics or principles which are used to manage the complexity in object-oriented programming are also considered to be the foundational and are generally accepted characteristics of object-oriented analysis, design, and programming.

5.1 Architecture of Energy Conservation Management System

Energy management is becoming increasingly important because of it role. However, developing systems that manage energy as a critical resource is not a new concept; there are many ideas in this sub domain. The architecture incorporates many concepts which include the classification of energy as first class resources, prioritizing resource request, accounting for fine-grained energy consumers, allocating resources based on dynamic constraints and providing quality-of-service.

For this project the mechanism to monitor and control system and component resource usage was adopted. The energy management architecture consists of a basic control device. The control device serves as the energy sources and output manager. The control device provides a set of control to the energy sources using the programming algorithms and communicates the energy source to the load and the battery bank. The control device component selects the input energy source via a set of conditions. It will monitor overall system energy such as the remaining battery energy and incoming harvested energy. The control device component uses energy information from the input sources to enforce that the specific load is powered by a specific source [34].

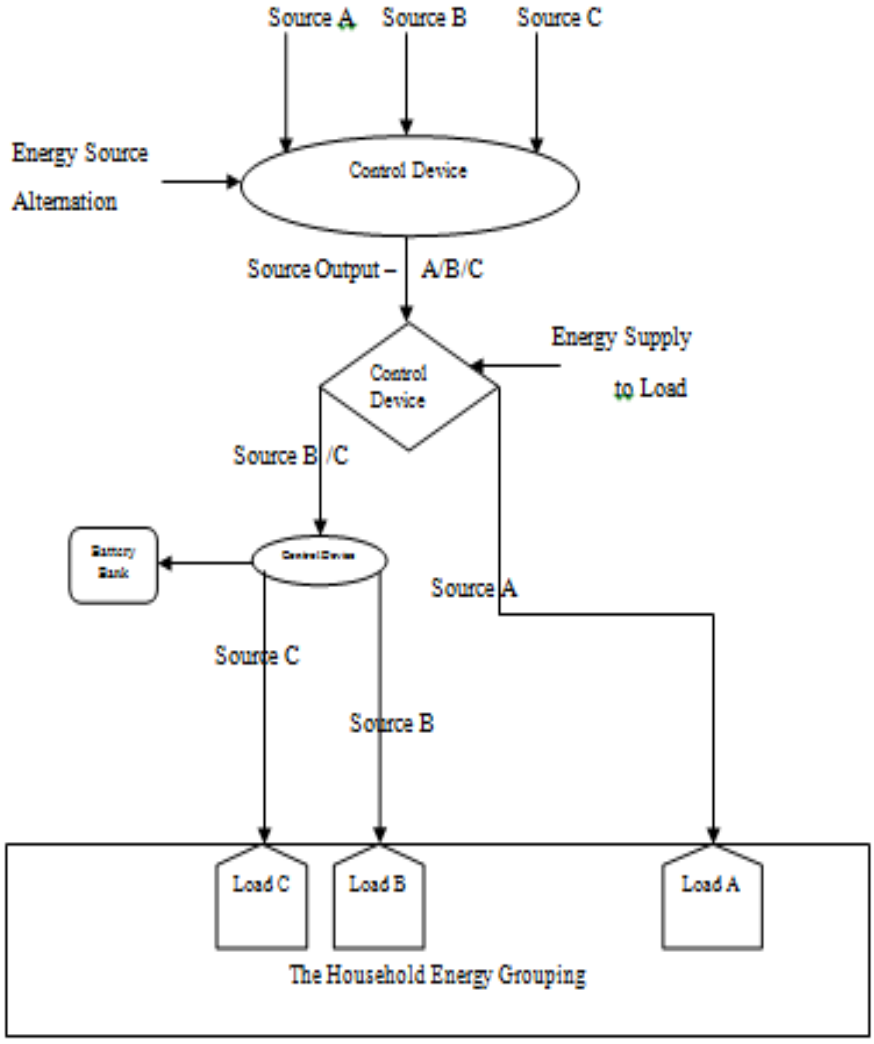


Fig. 1: Renewable Energy Management System Architecture

The control device component in the energy management systems provides an interface to the upper layer programming model. The programming, which is usually written in a high-level language, is responsible for translating the user requirements through the use of the ECMS. The programming that is used is MathLab. This is the way the system analysis and design of energy conservation management system will be done. The hybrid methodology will be employed. This will help to maximize the methodology's utility to the particular model of ECMS.

The management system includes the grouping of the household appliances and how they interact with power supply to conserve energy consumption through the distribution board [34]. The goal is to provide energy management solution that will harness the energy sources (renewable energy, fossil fuels) but not only that, a system that will optimize the utilization at the devices which will lead to energy conservation. This will minimize the energy waste without necessarily affecting the comfort level and quality of life. Hence, the reduction of environmental effects of production of energy will be minimized.

The model is important because it represents the rules that govern the properties of the organization of the system. The data model, during the requirements structuring phase, captures the conceptual data requirements for the conservation system. At the Logical Design phase, the data model will be refined. Overall, an effective conceptual model will be designed. The conceptual model of the ECMS will represent the outlook structure of the management system.

6. THE ECMS CONCEPTUAL MODEL

To achieve the energy conservation management system, it will compose of 10 stages. The system will be called energy conservation management system (ECMS). The first step will be to understand the needs of the users, requirement analysis, renewable technologies, materials, and energy integration and ends up with a comprehensive design plan for the user. The design phases of ECMS are in such a way that there are checkpoints. This will help the installer to check for the synchronization and conservation of energy issues of the renewable installation. The use of the proposed model to design the synchronization and conservation will reduce energy consumption and seamlessly integrate the energy sources. It is possible to improve on the model by carefully reviewing each part of the components. The details of ECMS are depicted in Figure 2 below followed by the Use Case in Figure 3.

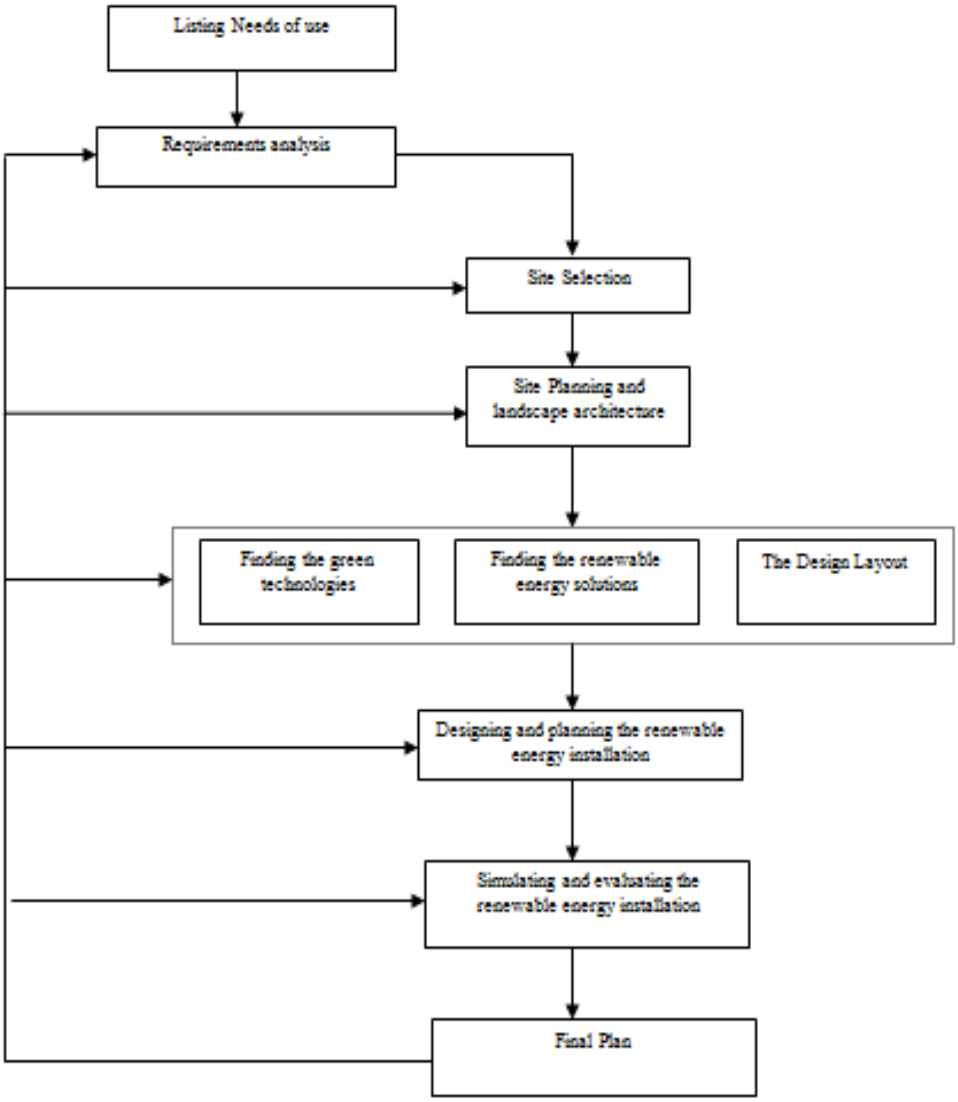


Fig. 2: The Renewable Energy Conservation Conceptual Model (ECMS)

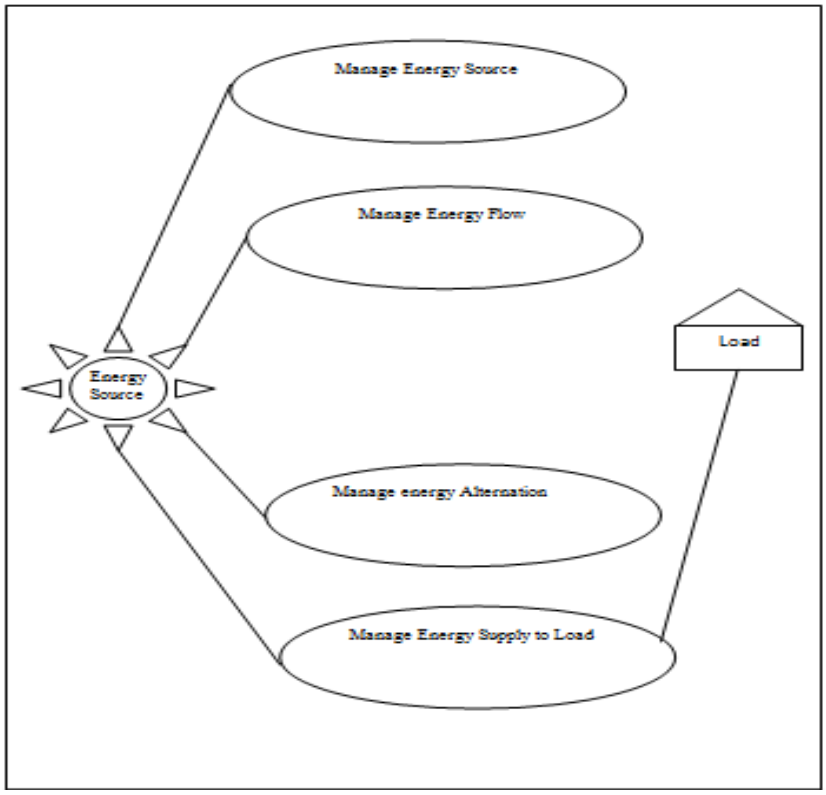


Figure 3: Use Case Diagram ECMS

7. CONCLUSION

If the prediction of the future of human population is anything to go by, then there could be a gloomy future for supply of energy and the human race. The supply has to be increased to meet the demand in the future. Unless there is an alternative way to meet the demand of energy, the energy will probably be a source of friction in human affairs. There should be a deliberate attempt to conserve electricity. The energy consumption can be reduced through effective conservation. The promotion of energy efficiency in each of the equipment will lead to overall energy conservation without necessarily reducing output of quality. The conceptual model for energy conservation in renewable energy is presented to help reduce the consumption of energy. The conceptual model when implemented will help to achieve the purpose of the conservation and synchronization. Saving of electricity and purposeful energy efficiency will lead to energy conservation. Energy management systems can be used to monitor and integrate energy –saving solutions which will lead to conservation of energy. These measures will help create a sustainable future for energy supply.

REFERENCES

[1] Jacobson, M. Z., & Delucchi. "A Path to Sustainable Energy by 2030", *ScientificAmerican, 5*(301), pp. 58-65, 2009.

[2] Pimentel, M. H. "Renewable Energy: Current and Potential Issues", *BioScience, 52*, pp.1111-1120, December, 2002.

[3] Leone, Steve. "U.N. Secretary-General: Renewables Can End Energy Poverty", *Renewable Energy World*, August, 2011.

[4] Diesendorf, M. *Greenhouse Solution with Sustainable Energy*. Sydney: University of New South Wales Press, 2007.

[5] Edenhofer, O. *Renewable Energy Sources and Climate Change Mitigation.* Cambridge: Cambridge University Press, 2011.

[6] Opatokun, S. A. Nassereldeen, A. K., Zahangir Alam, MD. & Shittu, A. J. K. Wastesoft- expert system for operational and integrated decision support system for solid waste management. *International Symposium on Sustainable Systems* and the *Environment* "ISSE'11". March 23-24, 2011. American University of Sharja. UAE, 2011.

[7] Jackson, R. D. "Public Opinion on Global Issues: Chapter 5b: world Opinion on Energy Security", *Council on Foreign Relations*, 2012.
[8] Goodall, C. *Ten Technologies to Fix Energy and Climate*. London: Profile Books, 2009.
[9] Agency, I. E. *Developing Renewables 2011: Best and Future Policy Practice*. OECD/IEA, 2011.
[10] Hirst, K. K. "The Discovery of Fire". Retrieved 15 August 2013, from www.about.com, January 2013.
[11] Brown, A. Marilyn and Socacool, Benjamin K. *Climate Change and Global Energy Security*. Massachussetts: MIT Press, 2011.
[12] Zubrin, R. *Energy Victor*, Prometheus: Books United States, 2007.
[13] Kovarik, Bill. "The Surprising History of Sustainable Energy", *The Source: Exploring the History of Sustainable Energy*, Retrieved 18 September, 2013 from http://sustainablehistory.wordpress.com, Posted on March 29, 2011.
[14] Jones, Geoffrey & Bouamane, Loubna. "Power from Sunshine": A Business History of Solar Energy", *Harvard Business School*, Working Paper 12-105, May 25, 2012.
[15] Schneider, Stephen H., Rosencranz, A., Niles, J. O., (eds.). *Climate Change Policy: A Survey, Island Press*, Wahington D. C., 2002.
[16] Evenden, R. *"Renewables Global Futures Report 2013"*. Sustainablehistory. Retrieved August 13, 2013, from www.wordpress.com, (2013).
[17] Brussels. "*Renewable Energy ProgressReport*", Report From The Commission to The European Parliament, The Council, The European Economic and Social Committee and The Committee of The Regions. Retrieved 08 September 2013 from http://ec.europa.eu/energy/renewables/reports/d /com_2013_0175_resen.pdf. Brussels, 27.3.2013. COM (2013) 175 final. {SWD(2013) 102 final}, 2013.
[18] REN21. "*Renewables Global Futures Report*", Renewable Energy Policy Network for the 21st Century. Retrieved 20 September, 2013 from http://www.cpuc.ca.gov/NR/rdonlyres/B185EAA-4A52-463C 9D6E6BD458F21A64/0/REN21_GFR_2013_pint.p df, 2013.
[19] Sambo, A. S. "*Matching Electricity Supply with Demand in Nigeria*", International Association for Energy Economics, Fourth Quarter, pp 32 – 38, 2008.
[20] Okoro, O. I., Madueme, T. C. "*Solar Energy Investments in a Developing Economy*",Renewable Energy, vol. 29, pp. 1599-1610, 2004.
[21] Cole, A.O. "*Restructuring the Electric Power Utility industry in Nigeria*", Proceeding, 20th National Conference of the Nigerian Society of Engineers (Electrical Division), October6-7, pp.1-6, 2004.
[22] Pimentel, David, et al., "*Renewable Energy: Current and Potential Issues*" Article December Pimentel, 1120 BioScience, December 2002 / Vol. 52 No. 12, pp 1111-1120, 2002.
[23] Jacobi, Patrick R. Renewable Portfolio Standard Generator Applicability Requirements: How States Can Stop Worrying and Learn to Love the Dormant Commerce Clause, 30 Vt. L. Rev. 1079, 1132, Summer, 2006.
[24] Bosselman, F. et al., *Energy, Economics and the Environment*, 3rd ed. Foundation Press: New York NY, 2010.
[25] [25] *"History of PV Solar"*, from www.Solarstartechnologies.com, Retrieved August 23, 2013.
[26] Christopher, A. Simon. *Alternative Energy Political, Economic, and Social Feasibility*, Lanham, Maryland: Rowman & Littlefield, 2006.
[27] Tsai, C., Su, P., Lin, C., Chang, K., & Chiang, M. "*A Conceptual Model for Green Building Design"*, *6*, 560–565, 2012.
[28] Girardet, H. *Surviving the Century: Facing Climate Chaos and Other Global Challenges*. Oxford: Earthscan, 2007.
[29] Johnson, T. R. "*The Xerx Star: A retrospective",* IEEE Computer. September, 1989.
[30] W. Wang, R. Zmeureanu and H. Rivard, "Applying multi-objective genetic algorithms in green building design optimization, Building and Environment", 40(11), pp.1512–1525, 2005.
[31] Olgyay, V. & Herdt, J. "The Application of Ecosystems Services Criteria for Green Building Assessment", Solar Energy. Science Direct, 26 February, 2004.
[32] Ekpo, E. Imo. Challenges of Hydropower Development in Nigeria. African Business, No. 323. Retrieved from http://www.energy.gov.ng. Magazine article, August –September 2006.
[33] Energy Commission of Nigeria. "Energy Resources Review", vol. 4, No. 3, pp.7-10, 2003.
[34] Henderson, J. J. Conceptual models. *Interactions*, pp. 25-32. January-February, 2002.
[35] Noon, R. *"Renewables global futures report 2013"*. Sustainable History. Retrieved August 13, from www.wordpress.com, 2013.
[36] Socacool, Benjamin K. and Brown, Marilyn A. *Energy and American Society - Thirteen Myths*. New York: Springer, 2007.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Prototype of Internet Model for Court Case Citation

O.M. Omisore*, O.C. Akinyokun, G.B. Iwasokun & O.W. Samuel
Department of Computer Science,
Federal University of Technology
Akure, Ondo State, Nigeria
E-mail: ootsorewilly@gmail.com*
Phone: +234-703-1967847

ABSTRACT

Citation of court cases stands a major problem in legal jurisdiction as cases are judged from accumulation of previous judgments using Rules of Precedent. Manual approaches adopted by citation of court cases in developing countries have led to long period between the time cases are reported and when accurate judgments are delivered. In this paper, the design of an Internet model for citation of court cases is proposed. The model adopts general relational database concept in an Internet environment. The model is implemented on Windows Vista Home Premium and WAMP APACHE as its web server. Case study of court cases handled in Ondo State High Court (OSHC), Akure between 2000 and 2012 was carried out to show the effective performance of the system.

Keywords: Web System, Case Citation, Court System, Relational Database Management System, Information Technology

African Journal of Computing & ICT Reference Format:
O.M. Omisore*, O.C. Akinyokun, G.B. Iwasokun & O.W. Samuel (2014). Prototype of Internet Model for Court Case Citation.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 77-84.

1. INTRODUCTION

Criminal activity continues to be a major concern in contemporary societies [6]. Most nations are faced with unacceptable levels of delinquency and crime, high crime rates is statistically normal feature of life all over the world. Research efforts seeking to explain the geographic variation in the rate of crime has been ongoing for more than 150 years [7]. Crime is the commission or omission of acts that violates the law and this is punishable under the law. Such punishments varies depending on the weight of crime committed and societal rule guiding such crime, punishments include fines, imprisonment, property forfeiture, removal from public office, disqualification from offices and/or death. The conception of the rule of law assumes that citizens have moral rights and duties with respect to one another and political rights against the state as a whole [1].

Law is asocial mediator that shapes the politics and economics of any society[5]. Law is defined as official body of rules and regulation that are generally found in constitutions, legislation and judicial opinions which are used to govern a society in order to control the behavior of the inhabitants of such society. Positive law is enforced upon the demand of individual citizens through courts or other judicial institutions so that moral and political rights are recognized [2]. The role of emotion in law is more controversial, yet, deeply rooted in the decision making process of the common law court [4].

The greatest asset of any nation is the welfare of her citizenry; each nation has a department within its Police Force which is charged with the responsibility of crime investigation [3], the criminal investigation department relies, primarily, on the information collected from complainants, witnesses and existing records of criminal cases in an attempt to investigate a case at hand. In the previous years, monitoring and citation of court cases in developing countries are done manually, this leads to long period between the time a case is reported and when a truthful judgment is delivered [8]. Citation of cases stands a major problem in legal jurisdiction as cases are judged from accumulation of previous judgments using *Rules of Precedent* [9].Reports on previous judgments are often kept piece meal in file cabinets in developing countries [8].

Hence citation of court cases, which require searching for cases, had posed a great deal of problem in the administration of Law and Justice. Also, there stands a chance of justice being switched unduly and known well to us is that *justice delayed is justice demand.* Hence there is need for speeding an accurate judgment for crimes cited at the court rooms. In early judicial tribunals, judges sat in enclosures, lawyers and the general public remained outside a bar [23]. Modern British courts are divided into those trying criminal cases and those trying civil cases; a second distinction is made between inferior courts, or courts of first instance and superior courts, or courts of appeal [11].

In the U.S. each state has its own system of courts, usually consisting of a superior (appellate) court, trial courts of general jurisdiction and specialized courts. The U.S. also has a system of federal courts, established to adjudicate distinctively national questions and cases not appropriately tried in state courts. A court is a form of tribunal, often a governmental institution, with the authority to adjudicate legal disputes between parties and carry out the administration of justice in civil, criminal and administrative matters in accordance with the rule of law [13]. In both common law and civil law legal systems, courts are the central means of dispute resolution and it is generally understood that all individuals have the right to bring their claims before a court, similarly, the accused have the right to present defense before the court. Court facilities range from simple and very small facilities in rural communities to large sophisticated technology in cities. A person learned in the law as an attorney, counsel solicitor and is as well practicing law is known as a **Lawyer** [11].

The role of the lawyer varies significantly across legal jurisdictions, the rules and regulations used to maintain the stability of political and social authority of a society and as well, to deliver justice becomes the role of specific lawyer [12]. Legal practice involves the practical application of abstract legal theories and knowledge to solve specific problems. This research presents a tool that can serve as a partner to legal practitioners so as to provide an optimal solution to the problems encountered while delivering case judgments in law courts and also to reduce or eliminate problems associated with existing orthodox systems. In the research, the design of a computer model for the citation of court cases is presented, the development of an Internet based system for citation of court cases is also presented and lastly, a case study of cases administered in Ondo State High Court (OSHC), Akure (Nigeria) between the years 2000and 2012 was implemented to show feasibility of our system. The paper is organized as follows: Section 2presents the research background and related works; Section 3discusses the system's architecture, conceptual design, and logic for query transactions. Section 4contains the implementation of the model in a web environment; section 5 presents the conclusion.

2. RESEARCH BACKGROUND

This section presents review of administration of Law and Justice in Nigeria, application of Information Technology (IT) in courtrooms, and study of court cases and their citations.

2.1 Administration of Law and Justice in Nigeria

Nigeria came to independence with a well-established legal system which includes court system and thriving legal profession in the British tradition [16]. The Nigerian legal system is based on English common law legal tradition by virtue of colonization and the attendant incidence [17]. Majority of Nigerians only encounter the legal system when they experience serious problems like arrest, fight, theft, breach of contract, divorce, homicide, or in cases of death where is need for probate to have a will approved or the property of the deceased distributed [18].

The Federal Republic of Nigeria operates two types of court systems which are superior and inferior courts. Superior courts refer to all the courts presided over by judges trained in law and dutiful to recording and publishing statutory for public access proceedings. Decisions made in superior courts are binding on all other courts. Supreme Court and Court of Appeal are examples of superior courts. Inferior courts may not have legal practitioners as presiding officers and neither is its decisions binding on any court nor is the decision made there are reported in law reports. Tribunals and Magistrate courts are examples of inferior courts [17]. The courts available for the administration of law and justice in Nigeria are conceptually shown in Fig. 1based on their hierarchies.

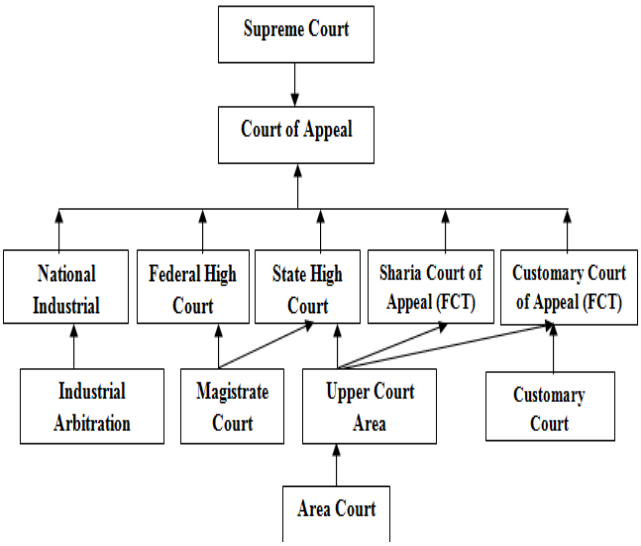


Fig. 1 Hierarchy of Nigerian Courts

The Supreme Court which is at the top of the hierarchy has final and overall power over the decisions earlier taken by other courts, and next to it is the Appeal Court, while Area Courts are found at the tail of the structure.

2.2 Literature Review of IT in Courtrooms

Experts Systems (ES) are designed to solve complex problems by reasoning about knowledge, like a human expert, and not by following the procedure of a developer as is the case of conventional programming [21] described ES as computer based programs that stores expert's knowledge and utilize it in deductions and inductions. Traditional approach of wading through a host of books and other physical documents is fraught and disadvantageous, there is a need for legal research to adapt to the changes and provide us with inventions of IT.

The procedures of how Information System can automatically receive signed case file documents via electronic mail are discussed in [22]. The information system also checks each document for conformity to court requirements. Reference [23] accessed the influence of Computer Aided Legal Research (CALR) with regards to determining its effects on the qualities and types of legal authorities cited in courtrooms. CALR is the technology of using computer system in the citations of court cases. Observations on the output of courtrooms had shown that there is an apparent change in legal authorities cited in law courts, no legal evidence could attributed to the advent of CALR [25], furthermore, [23] stood CALR out clearly to be a better direction in the practice of legal issues and authority citations. Judges, of course, consult legal precedent when making their decisions and citation to legal authorities in support of their holdings. [26] presents the effects of technological advances in courtroom.

The pilot program launched in the administrative office of the United States courts funded advanced technology that facilitated US courtrooms. The proliferation of technology in the courtroom does not only affect the processes of court trials but, also changes the way law is been practiced [24]. Most people will never face a felony trial, but many are summonsed to court on lesser charges, there is need to guide people's sense of general knowledge of procedural fairness in legal system. Defendants fail to show up in court for relatively minor offences due to their nature and this easily leads to long period between the citation and court date. Claim of Failure to Appear (FTA) is expensive mostly on maintenance; such cost extends to victims, witnesses and even the defendants. Ref. [7] proposed a model with which court appearance rates can be increased using live-caller telephone court date reminder system.

2.3 Application IT in Court Case Citations

Court case is a dispute between opposing parties over a particular matter and such are usually resolved by courts, or by some equivalent legal process [Wikipedia, 2012]. Court case management is a subset of law that covers the approach and technology used to leverage knowledge for managing the life cycle of a case more effectively [10]. Court case management has two primary functions, these are to manage, consolidate, share, and protect information, and to track and shape the business process.

Citation of court cases is a major factor in the administration of Law and Justice. This is regarded as the extraction of legal authorities from legal material known as Law Reports or in a 'neutral' form which will identify a decision wherever it was reported. Law reports are a series of reports of cases decided by the superior courts and other courts in the past which can serve as templates when deciding future cases[9]. The basic assumption underlying legal proceedings is that the procedures that had influenced the decision of a court in the past and present will continue to do so in the same manner if the facts constituting the later are similar to those of the former. Hence, the major goal of law reports is to identify and isolate those facts of a court in a given case.

Case citations can be formatted in different jurisdictions, but generally, they contain the title of the reports, volume number, page number, and year of decision. Most decisions of courts are not published in printed law reports. The expense of typesetting and publishing them has limited the printed law reports to significant cases. Internet publishing of court decisions resulted in a flood of information [9]. The task of a lawyer while attending to cases in the courts of law is to resolve the facts of the case brought to him into their legal categories and then search for authorities to prove his points. Location of an authority in a law report can be achieved through direct searching through law reports or searching through law indexes. In the direct searching through the law reports, a legal practitioner in a case picks each of the numerous volumes of the law reports, one after the other, and check through the index of authorities that is contained at the back page of each of the publication. If a particular volume does not contain the destined authority, he or she drops it and then pick on another volume. The searching continues until the desired authority is located.

In reference [7], a model for analogy and decision making in the citation of court cases was proposed using soft computing tools. [14] presents a computer aided system for monitoring and citation of court cases for standalone workstations and those in Land Area Networks, observing a case study of OSHC, Nigeria. [14] examined the current situation of delay in the federal districts court of U. S.

3. SYSTEM ARCHITECTURE

The proposed architecture for citation of court cases in law courts is presented in Fig. 2.

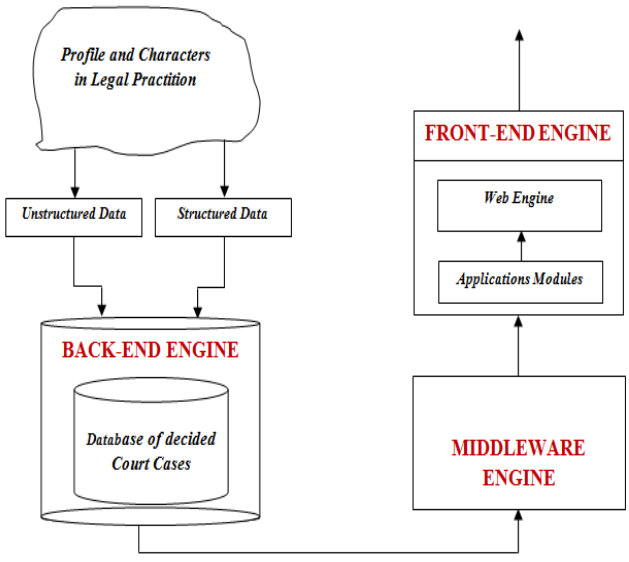


Fig. 2 Architecture of Court Case Citation Model

The architecture is composed of three main components which are:

A. *Back-End Engine.* This manages the data required by the application. This includes the details of profile and characters involved in legal practice, and the details of cases decided in some law courts.
B. *Middleware Engine.* It does the actual processes for a chosen transaction so to cite relevant case; and
C. *Front-End Engine.* This is composed of the Application Modules that enable users to make use of the data stored in the database of the system and Web Engine which aids creation and management of system data in an online and real-time mode.

The citation of court cases database is modeled as a network of semantically related objects (files). The relational model proposed by [19] for database system, with the form $R[a_1, a_2, \ldots, a_{k+1}, a_k, a_{n-1}, a_n]$ is adopted in this research. R is the name of the relation while set $\{a_i\}$ represents its decision variables. The database schema used for the web based system has 11 inter-related files for keeping dependent information. These files and the relational flow are shown in Fig. 3. The Internet based model for court case citations is designed to run in a web environment using the architecture displayed in Fig. 2 above as the framework. The web based system is developed on a three tier mode featuring: *Front-Tier, Middle-Tier, and Back-Tier*.

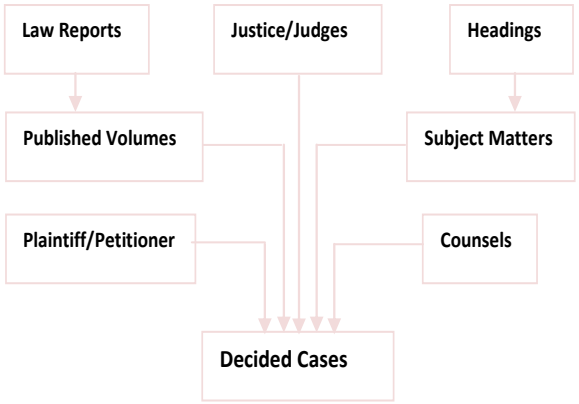


Fig. 3 Database Files for Court Case Citations

The **front tier** can be described as the presentation layer made up of series of web pages accessible by legal practitioners with the aid of a web browser. Some web pages are purely informative and do not require usersauthentication, while others explicitly authenticate the presence of users. This tier presents a list of pages, starting with *index page*, to users. Index page is the first page displayed upon starting the system. It presents the vision and mission of the system, links and menu options to other pages, and some other information. The **middle tier** is the component of the web system that contains the logic of data processing. It makes decision such as which cases are to be cited, which content is filled onto a page template, and the list of functions that are available to a user. It performs the actual authentication and authorization of the application users. The middle-tier also serves as a bridge between the front tier and back tier. It performs a list of manipulations as required by the system, and how the information derived from processing the data is summarized and reported to the outside world.

The **end tier** is the database system where the on which the web system relies functions is resided. There are several database systems that support web programming and data processing in a web environment. However, the choice of database system is the prerogative of the database administrator. The users' activities on the database are termed transaction. Two major types of transaction considered in this research are Update and Query transaction. Update transactions can only be carried out by specific users as this contains exclusive operations. Query transactions are utilized to question the database so as to generate useful reports. Typical query transactions are:

i. For a given type of crime, such as 'Murder', cite a list of related cases.
ii. In the 'Supreme Court', cite the cases handled in between January 2009 and December 2011.
iii. Given a Justice or Judge, cite the cases he/she presided over.
iv. Given the name of a Lawyer, cite all the decided cases that he/she handled in the Supreme Court.
v. Report on all court cases before the Supreme Court that were on appeal from Federal Court of

The transactions support declarative processing that isa user is only expected to know what to do but does not need to care about how it will be done. The program plays the role of an interface which carries the granular or detailed codes of how to produce reports to the commands issued by the system user. This facilitates an intelligent, user friendly and interactive system. The varying categories of users of the application are system expert users, system end users, and system casual users.

4. SYSTEM IMPLEMENTATION

To test the practicability of the formulated model, the system was implemented as a three tier web system with front engine, middle engine, and back engine.HTML tags were used to structure the templates of each page while Java script determines their behavior; PHP scripting was used to coordinate the activities of the system at the middle end, while MySQL DBMS was taken as the back end engine. The system knowledge is stored and manipulated at the back end. The web based system is developed to run either on Localhost characterized by **W**indows **A**pache **M**ySQL **P**HP (WAMP) Server or Internet with a view to ensuring online and real time access to relevant knowledge on previously decided court cases in Nigeria. The former is adopted in this research. The activation of the system produces a splash screen with login form shown in Fig. 4.

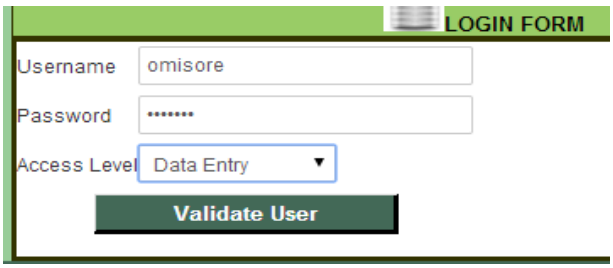


Fig. 4 Users Authentication Login Form

The user is prompted to supply his login details. The categories of users considered in this implementation are: *Administrator*, *Data Entry*, *Others*; each category has distinct privilege level.

The available privileges are: *full privilege*, *creation-update privilege*, and *read-only privilege*. Authentication is done upon supplying a valid login details and if successful the home page is displayed with the menu bar activated. The home page presents links to other modules of the system. Other modules are: Entry Module, Citation Module, Update Module, Transaction Module, and Report Module.

Entry Module offers administrator users the ability to create new records to any system file. The system files include: *Cases*, *Judges*, *Counsels*, *Litigants*, *Citations*, *Courts*, *Headings*, *Subject Matters*, *Law Reports*.

Case Citation Module contains series of web files used to cite cases on demand. Citations are done by Boolean retrieval technique with proximity operators and wildcards. Cases are either retrieved by alphabetical search or chronological indexes. A view of the form used for capturing case details is presented as Fig. 5.

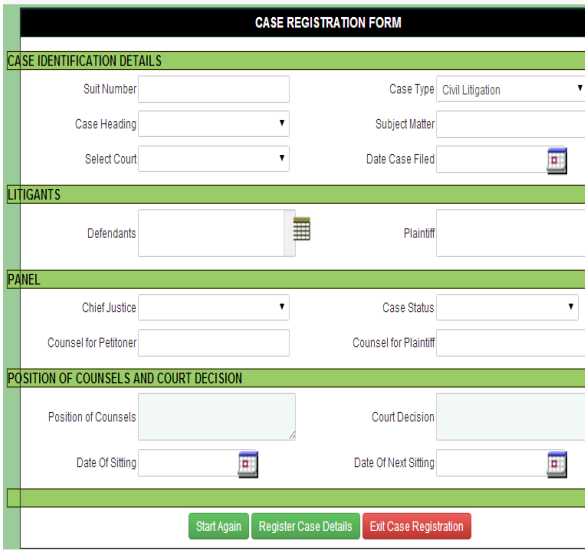


Fig. 5 Entry Form for Creating Court Cases

Update Module has two sub options which are record modification and deletion. Both modification of existing records and deletion of unwanted records in the system files are done in this module to aid database's integrity. *Transaction Module* contains pre-codified query transactions, T1, T2,..., Tn. Selection of any transaction, say, T1 executes suite of programs with Tag 'T1'and its success cites a list of cases with options to view or print.

Report Module is to generate reports for mined information. A click on any sub menu option of the module shows all records in the system file attached to the option selected. It is noted that all users have the privilege to access this module.

5. CASE STUDY

Case study of crimes handled at OSHC, Akure was used to test the system; record of major crimes administered by the court between the years 2000 and 2012 was gotten during visitations to the court. The crime records extracted are:

a. Forgery and Stealing;
b. Man Slaughter;
c. Armed Robbery;
d. Rape and Kidnap;
e. Conspiracy; and
f. Murder.

This is tabulated below as Table I. Some of the analysis drawn from statistical records of the tables shows that:

a) Murder cases administered was at minimal level in 2009 and later keeps increasing;
b) Stealing and forgery cases handled between the time range is at a uniform level;
c) Armed robbery cases recorded in 2010 has the highest record in the time range;
d) Number of crimes committed in 2009 is the lowest while highest number of crime was committed in the year 2012.
e) Record of cases administered in the years 2000, and 2009 were poorly documented;

Table 1: Statistics of Cases Logged at OSHC, Nigeria, between January 2000 and December, 2012.

	a	b	c	d	e	f	Total	%
2000	2	0	0	0	2	4	**8**	**1.8**
2001	1	2	0	0	0	0	**3**	**0.7**
2002	3	0	5	0	1	0	**9**	**2.0**
2003	0	0	0	0	0	0	**0**	**0**
2004	4	6	26	3	3	10	**52**	**11.7**
2005	3	13	34	3	0	15	**69**	**15.6**
2006	5	1	25	0	2	7	**40**	**9.1**
2007	6	2	15	2	7	6	**38**	**8.6**
2008	3	1	4	0	3	1	**12**	**2.7**
2009	1	0	0	0	1	0	**2**	**0.5**
2010	5	5	36	0	16	12	**74**	**16.7**
2011	2	0	21	1	7	6	**37**	**8.4**
2012	3	4	24	3	44	20	**98**	**22.2**
Total	**41**	**36**	**192**	**12**	**81**	**81**	**442**	**100**

Data adapted from the OSHC Registry on 12th of February, 2013.

Fig. 6 shows, with the aid of pie chart, the annual statistics of cases recorded by OSHC, between January 2012 and December 2012.

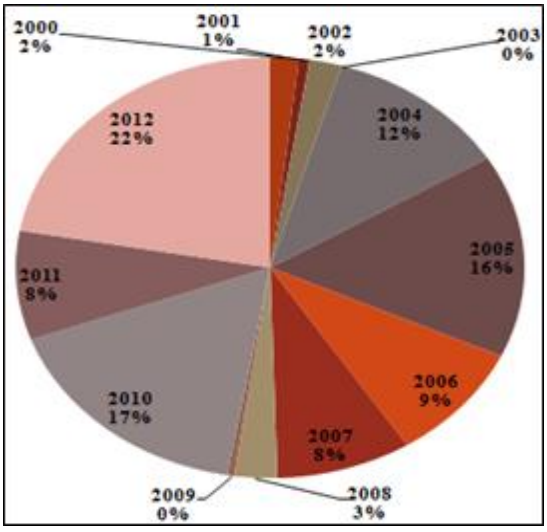


Fig. 6 Annual Statistics of Cases Recorded in the Study Timeline.

6. CONCLUSIONS

In, this paper an attempt was made to develop a web based system for the citation of Court Cases. Web based citation system gives room for global maintenance of legal documents in an efficient manner over the orthodox methods and therefore aids a greater amount of citation within a favorable time range. This work presents an Internet model with which previous court decisions can be located and cited in minimized time and as also how such decisions are accessible in an online and real-time manner. The main objective of the research is to develop a web based system for citation of court cases which can be augmented with the orthodox methods and not meant to replace orthodox systems. It is clearly shown that the deployment of IT facilities in legal practices had drastically reduced the stress passed through while adjudging a matter and had brought a better result.

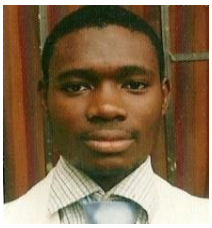
Area of future researches includes the application of Text-to-speech in Law and Justice. Other security approaches like biometric system and graphical password system can be adopted as security mechanisms for court case citation systems. Areas of application of ES techniques such as Neural Networks and Fuzzy Logic in court cases citation are still lacking.

Acknowledgment
We wish to acknowledge the staff of OSHC, Akure, Nigeria, for their assistance in the area of data collection and analysis, and the chambers of O. J. Jejelola, and Michael Owolabi for their contributions.

REFERENCES

[1] D. Ronald, "Political Judges & Rule of Law in a Matter of Principle", Harvard University Press, Cambridge, 1985.
[2] S. Rosemary, "Law Students & Hypertext: One Law School's Model", BILETA 1996 Conference Proceedings, 3 The Journal of Information, Law and Technology,pg 1-14, 1996.
[3] O. Akinyokun and O. Olabode "Framework for Client-Server Distributed Database System for Population Census", Pacific Journal of Science and Technology, Vol 9, No. 2, 2008.
[4] D. David andL. Macfie, "The Theory of Moral Sentiments", Oxford University Press, eds., 1975
[5] A. Norrie, "Crime Reason and History", Cambridge University Press, London, 2001, Pg 199-210, 2001.
[6] W. Ackerman and A. Murray, "Assessing Spatial Patterns of Crime in Lima, Ohio", Ohio State University Journal, Vol. 21, No. 5, Pg. 423-437, by Elsevier Ltd, USA.
[7] H. Jurgen, "Analogy Making in Legal Reasoning with Neural Networks and Fuzzy Logic", Artificial Intelligence & Law, vol. 7, Pg. 289-301, Kluwer Academic Publishers, Netherlands, 1999
[8] O. Akinyokun, "Computer Aided System for Monitoring and Citation of Court Cases", Law and Practice Journal; Published by the Nigerian Law School, Bwari, Abuja, Nigeria, 2006.
[9] W. Glanville, "Learning the law", 14th edition, Sweet & Maxwell Ltd, pp 4-15, 2010
[10] Sanni, A. O. (2004), "Introduction to Nigerian Legal Method", Kuntel Publishing House, Ile-Ife, Pg 4-10
[11] C. Geoffrey and D. Angelo, "Legal Ethics - A Comparative Study", Stanford University Press, ISBN 0-8047-4882-9, Pg. 20-23, 2004.
[12] H. John and P. Rogelio, "The Civil Law Tradition: An Introduction to the Legal Systems of Europe and Latin America", 3rd edition, Stanford University Press, Pg. 102-3, 2007.
[13] D. Walker, "The Oxford companion to law", Oxford University Press, Pg. 301, 1980.
[14] L. Mattia, "Justice Delayed - An Overview of the Options to Speed Up Federal Justice", Journal of Public and International Affairs, Woodrow Wilson School, Vol. 18, No.6 pg. 127-149, 2007.
[15] J. Ohireime and A. Asein, "Introduction to Nigerian Legal System", Ibadan, Sam Bookman Publishers, 1998.
[16] M. Oknay, "Nigerian Commercial Law", Revised Edition (2009), Africana First Publisher PLC, ISBN: 978-978-175-629-0.
[17] A. Mojeed and O. Soyeju, "Courts and Rudiments of Nigerian Law", African Collectives, Malthouse, Press, Nigeria, 2002.
[18] S. Savory, "Artificial Intelligence and Expert System", Chichester, England, pp. 28-29, 1998.
[19] E. Codd, "Relational Model of Data for Large Shared Data Banks", Communication of ACM, 13 (6), pp. 377, 1970.
[20] A. Alan, "Electronic Filing and Judicial Electronic Document and Data Interchange (JEDDI)", Proceedings of the Fourth National Court Technology Conference (CTC4) of the National Centre for State Courts, 2002.
[21] P. Hellyer, "Assessing the Influence of Computer-Assisted Legal Research", Law Library Journal, San Jose University, Vol. 97, No 2, Pg. 285-298, 2005.
[22] J. R. Hamilton, Comment, Computer-Assisted Legal Research, 51 OR. L. REV. 665 - 676, 1972.
[23] D. Kuchler and C. Leslie, "How Technological Advances In The Courtroom Are Changing The Way We Litigate", Federation of Defense And Corporate Counsel Journal, FDCC Quarterly, Vol. 58, No. 2, Pg. 241-255, 2008.
[24] P. Michael and H. William, "Trial Presentation Technology: A Practical Perspective", Tenn. L. Rev, 587, 2000.
[25] R. Timothy, R. Michael, and M. Dorian, "Increasing Court Appearance Rates and Other Benefits of Live-Caller Telephone Court Date Reminders", Journal of the American Judges Association, American Judges Association, 2011.
[26] U.S. Courts, "About Case Management/Electronic Case Files (CM/ECF)", July 2009.
[27] D. Spencer and A. Foster, "Capturing More Time and Keeping Your Clients Happy While Doing It", Law Practice Today, USA, 2007.

Authors' Biography



Mr. Omisore Mumini Olatunji is a research student at the Department of Computer Science, Federal University of Technology, Akure, Nigeria, where he obtained a Bachelor of Technology Degree in Computer Science in 2009 and a Master Technology Degree in the same department in 2014. He has interests in Software Engineering, Computational Intelligence, Digital Libraries with specialties in Data and Knowledge Mining. He is a member of Institute of Electrical Electronic Engineers (IEEE), Institute of Electrical Electronic Engineers Computer Society (IEEECS), and Association for Computing Machinery (ACM) He worked as a Senior System Engineer at the High Technology Research and Development Group Computer Limited, Akure, Nigeria for four years and he has being actively involved in National and International research works with a good number of scientific papers published in reputable journals worldwide.

Professor Akinyokun Oluwole Charles had BSc. First Class in Computer Science and Mathematics from Obafemi Awolowo University, Ile-Ife, Nigeria in 1979 and PhD in Computing Studies from the University of East Anglia, Norwich, United Kingdom in 1984. He has specialization in Distributed Database System and serves as a Professor of Software Engineering at the Department of Computer Science, Federal University of Technology, Akure, Nigeria. He is a Fellow of Nigerian Computer Society (NCS), Member of the Computer Professional Registration Council of Nigeria (CPN), British Computer Society (BCS) and Association for Computing Machinery (ACM). He has been actively involved in Research, Innovation and Development (RID) in Knowledge Engineering in computer network environment since 1985. He has featured prominently as a Reviewer to many reputable Journals and Conference Proceedings, External Examiner and Assessor to many Nigerian and overseas universities. His scholarly publications have featured prominently in many reputable national and international conference proceedings and journals.

Dr. Iwasokun Gabriel Babatunde is Lecturer at the Department of Computer Science, Federal University of Technology, Akure, Nigeria. He obtained a Bachelor of Technology Degree in Computer Science in 1997, a Master of Technology Degree in 2006 and a PhD Degree in Computer Science from, Federal University of Technology, Akure, Nigeria. His research has focused on database systems, Web 2.0 and computer security with specialties on Biometric Systems. He has being actively involved in Research and has published reputable scientific papers in both national and international journals and conference proceedings. He is a Member of Nigerian Computer Society (NCS) and Computer Professional Registration Council of Nigeria (CPN).

Mr. Samuel Oluwarotimi Williams has B.Sc. Degree in Computer Science from Kogi State University, Anyigba, Nigeria and a Maters of Technology Degree in Computer Science at the Federal University of Technology, Akure, Nigeria. He has over five years' experience in teaching and research in the field of computing with special interests in Computational Intelligence, Cloud Computing, and Computer Programming. He currently works with State Information Technology Agency, Ondo State, Nigeria as a Programme Analyst. He is a certified associate at the Oracle University and he has professional Member of Association for Computing Machinery (ACM), Institute of Electrical Electronic Engineers (IEEE), IEEE Computer Society (IEEECS), and Nigerian Institute of Management (NIM). He has published a number of scientific papers in respectable international journals.

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

Evaluation of Unshielded Twisted-Pair Cable for Wired Data Networks

T. Oguntunde & A.O. Osofisan
Department Computer Science
University of Ibadan
Ibadan, Nigeria
tantos557@yahoo.com nikeosofisan@gmail.com

ABSTRACT
Generally, the sight of numerous kinds of cable in the Nigerian IT (Information Technology) market requires immediate alerted attention. The focus of this research paper is to look into the quality of collected samples of Unshielded Twisted Pair (UTP) cable available in Nigeria market by measuring some parameters or properties. Cables may look very similar on the outside, however, they can be completely different on the inside. Construction practices, materials used, quality control and other factors all contribute to the performance of a cable. The paper first renders a brief introduction on Unshielded Twisted Pair (UTP) cable. It then justifies the reasons why UTP, especially category 6 becomes the focus of attention out of various other UTP categories and other variety of other Network Transmission Media (cables through which electronic signals travel between a point A and another point B). The research paper also looks into the ideal qualities of a good copper-made cable for video, audio and computers. It finally presents a discussion on the analyzed parameters of collected cable samples.

Keywords: Quality, Transmission media, Determinant, Network throughput, Nigeria, UTP Cable Market Study

African Journal of Computing & ICT Reference Format:
T. Oguntunde & A.O. Osofisan (2014). Evaluation of Unshielded Twisted-Pair Cable for Wired Data Networks.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 85-90.

1. INTRODUCTION

Unshielded Twisted Pair (UTP) cable is the most populous amongst all Network Transmission Media. UTP is made of eight (8) copper wires intertwined into four (4) pairs. Comprehensive Cable [1]) established the fact that, Copper is the material of choice because of its conductive properties and flexibility which is important for cable durability. Copper is one of the basic chemical elements. In its nearly pure state, copper is a reddish-orange metal known for its high thermal and electrical conductivity [2]. Also [2] continued to say that, copper is combined with zinc to produce brass and with tin to produce bronze which improves the bandwidth (the highest frequency to which a cable will perform) and capacitance. Copper was first used as early as 10,000 years ago. In the United States, the first copper mine was opened in Branby, Connecticut, in 1705 followed by one in Lancaster, Pennsylvania, in 1732.

Pure copper is rarely found in nature but is usually combined with other chemicals in the form of copper ores. Cavette [2] exposited that, there are about 15 copper ores mined commercially in 40 countries around the world. The most common are known as sulfide ores in which the copper is chemically bonded with sulfur. Others are known as oxide ores, carbonate ores or mixed ores depending on the chemicals present. Many copper ores also contain significant quantities of gold, silver, nickel and other valuable metals, as well as large quantities of commercially useless material, he added.

2. LITERATURE REVIEW

Although UTP was once considered to be slower at transmitting data than other types of cable, this is no longer true. In fact, UTP is considered the fastest copper-based medium today, declared Cisco Systems (2003). Moreover, Gallen [10] quoted Kish [14]) as saying that, "From a performance perspective, Category 6 cabling provides twice the Bandwidth (200 MHz) and 16 times (12 dB) better than signal-to-noise margins compared with Category 5e cabling". Also, While cables may look very similar from the outside, they can be completely different on the inside. Construction practices, materials used, quality control and other factors all contribute to the performance of a cable [1] .

First, cable quality cannot be underestimated in successful networking because it is the medium through which data packets flow from point A to another point B on the network. Jay [12] confirmed this by saying that, the networking environment is determined by the quality of the cable, its length, patch panels, installation practices and local electrical noise. He explained further that, one thing hampering the migration to 1000Base-T is performance problems associated with the physical layer. The Physical layer defines the electrical, mechanical and functional specifications for establishing the physical links between systems. Moreover the Physical Layer is made up of the cable, connectors, the physical links on the devices and the encoding/ decoding circuitry of the network devices.

So, cable(s) is an issue of serious concern that is worth study especially in an open IT market like Nigeria's. Also Griffith et al [13] stated that, "However, in our experience, designers under-appreciate how strongly cable properties affect devices behavior in both experiment and simulation". Furthermore, many a times, manufacturers of cables introduce errors (intentionally or unintentionally) in the manufacture of cables. Common cable modeling errors include incorrectly determined electrical-parameter values, using lumped element representations where distributed-element representation are needed and even neglecting the cable entirely, Griffith et al [13]. This is also confirmed by [1], One thing to watch out for is some China factories and Internet importers have begun substituting a material known as copper coated steel for copper (pure). This is inferior conductive material and is not suited for audio/ video use.

Copper clad steel has much poorer and does not perform nearly as well". It is used by some companies because it is much cheaper and significantly lowers the cost of the cable as a whole [1]. Some manufacturers even reduce the wire guage (size) but trade-off performance to cut cost. The larger the size of the center conductor, the higher conductivity it will have because the larger the guage, the lower the resistance. The lower the resistance, the less signal loss a cable will have, the more signal that makes it to the display from the source, declared by [1]. Moreover, aside of the low cost of UTP compared with the coaxial cable and optical fiber cable, its flexibility and 10 Gigabyte throughput, UTP is also finding increasing use in video applications, primarily in security cameras. Many middle to high-end cameras include a UTP output with setscrew terminals. This is made possible by the fact that UTP cable bandwidth has improved to match the baseband of television signals, Chris Shore.

Table-1: The pre-analysis information of each UTP cable sample.

S/N	Cable name	Category	Guage	Standard
1.	A	6E	Not specified	ISO/IEC/11801
2.	B	6	23 AWG	EIA/TIA 568B.2/ISO/IEC 11801
3.	C	6	Not specified	Not specified
4.	D	6E	24 AWG	Not specified
5.	E	6	23 AWG	Not specified
6.	F	6	23 AWG	EIA/TIA 568B.2-ISO/IEC 11801
7.	G	6	23 AWG	EIA/TIA 568B.2
8.	H	6E	23 AWG	Not specified
9.	I	6	23 AWG	ISO/IEC 11801

This research work is testing for quality of UTP cable based on three parameters. Cabling performance parameters are not mysterious. They are symptoms that enable you to track down flaws that can distort and ruin the signal on twisted-pair cable. Flaws in the cable structure might be causing part of the signal to be reflected back to the source, resulting in problems there [7][8][15]. Expressing performance characteristics as measurable quantities makes it possible to set cable quality standards and use cable-testing equipment to detect faults". Moreover, according to [6] and [4] "Manufacturers know that a cable pair's impedance varies very slightly along its length. This is due to small variations in manufacturing such as concentricity, diameter, quality of twisting and wall thicknesses".

According to Hitachi Cable Manchester [11]in the past, shortages of some materials, including those used in making plenum rated cables forced manufacturers to find alternative compounds and alternative construction methods that would allow them to continue manufacturing and to pass the appropriate UL burn tests required for plenum rated cables. These compounds have a direct impact on the speed at which a signal will travel down the conductor (Nominal Velocity of Propagation), Hitachi [11] reiterated. Impurities within the copper, including oxygen, silver, iron, sulphur, antimony, aluminum and arsenic, coalesce at the grain surface, or boundary. This creates much higher impedance to the electron flow by essentially forcing the electrons to have to 'jump' those poorly conducting boundaries where grains touch". Reducing the impurity content and the number of grains per foot can make a wire a much more efficient conductor [9][14] .

3. METHODOLOGY

1. An equal length (2 inches) of each sample cable was cut
2. The jacket of the cut cable were stripped exposing the insulated 4-pairs of UTP
3. The 4-pairs were unwound giving eight insulated wires
4. The individual wires were stripped of their rubber insulation leaving bare supposedly copper wires
5. The eight copper wire pieces were all crunched together in a cup each. The cups were arranged in circular form on a tray of Minipal 4 Spectrometer (Energy Dispensing X-ray Flourescent)
6. Steps 1 to 5 was observed for each cable sample
7. Elemental Analysis was performed on the stripped copper wires in the different cups on the tray.

The findings are as displayed in table-2 below.

Table-2: Elemental Analysis of 9-UTP cable samples

Cable Name	Constituent makeup (%)												
	Si	Ca	Mn	Fe	Cu	Os	Al	K	Ni	Br	Yb	S	Mo
A	16	1.6	0.49	58.9	**21.3**	2.0	-	-	-	-	-	-	-
B	13	1.4		0.58	**85.0**	-	-	-	-	-	-	-	-
C	12	0.64	-	14.2	**19.8**	-	53	0	0.38	0.05	0.6	-	-
D	9.6	0.76	-	18.9	**33.5**	-	37	-	-	0.05	-	-	-
E	12	1.2		19.8	**66.3**	-	-	0.3	-	-	-	-	-
F	20	1.8	-	7.39	**51.9**	-	18	-	-	0.2	-	-	-
G	13	1.1	-	0.42	**73.1**	-	13	-	-	0.1	-	-	-
H	-	-	-	-	**6.7**	-	88	-	3	0.4	2	-	-
I	14	0.97	-	0.51	**70.1**	-	13	-	-	0.1	-	1	0.2

4. RESULTS AND DISCUSSION

UTP cable is supposed to be made of mainly pure copper but if electroplated with tin, it improves the bandwidth and capacitance, Chris [2]. In none of these cable samples is tin or silver used to improve cable performance. Looking into Table-2, most of these cable analyzed are alloyed with some other elements such as Si (Silica), Al (Aluminium), Bromide (Br) and others. It is only in cables B, E, F, G & I that copper (Cu) dominated. In cable A, Fe (Iron) dominated (58.9%) while in cables C, D & H, Aluminium dominated with 53%, 37% and 88% respectively. From this analysis, it has shown that most of the cables shipped into Nigeria might not perform optimally on the network when even the dominant compounds of manufacture are not copper. This was corroborated by eHow Tech(2009-2012) that, "If a difference in speed or performance is noticed in Ethernet cables less than 100 meters, the problem is probably with the quality of the cable itself, not the length". "Government regulations require that cable networks meet certain minimum standards for analog television signal CNR (Carrier-to-Noise Ratio) which is the difference between the amplitude of an RF signal and the amplitude of noise present in the transmission path of the RF signal" [4].

Nigeria should also enact such a regulation especially, now that she is launching a network-driven economy such as cashless Lagos (as a kick-off point) and various other e-payment schemes, network cable quality should not be compromised.

4.1 Number of Twists per Meter (Twists rate)

The twist rate (also called pitch of the twist, usually defined in twists per meter) makes up part of the specification for a given type of cable. Wikipedia continued to say that, where nearby pairs have equal twist rates, the same conductors of the different pairs may repeatedly lie next to each other, partially undoing the benefits of differential mode. For this reason, it is commonly specified that at least for cables containing small numbers of pairs, the twist rates must differ. Pairs having the same twist rate can still experience some degree of crosstalk, the site concluded. Diaverty [6] noted that, "When a signal travels down a conductor, an electric field is created, which interferes with any wire close by. Diaverty continued to say that, "This is Crosstalk and gets larger at higher frequencies and the more parallel the wires. The twists in the pairs should (in theory) cancel this effect. For good signal cancellation it is important that the twists are symmetrical and that adjacent pairs have different twists".

Table-4: Twists rate count per pair of the sample cables

Cable Name	Pair-1 (wbr/br)	Pair-2 (wbl/bl)	Pair-3 (wo/o)	Pair-4 (wgr/gr)
A	42	49	58	51
B	93	51	69	59
C	35	35	36	36
D	35	38	51	46
E	40	52	55	46
F	35	36	35	35
G	54	80	59	70
H	83	64	94	72
I	78	58	68	53

Where W/Br= White Brown, Br=Brown, W/Bl=White Blue, Bl=Blue, W/O=White Orange, O=Orange, W/Gr=White Green, G=Green and stand for the color-code of the UTP sample cable

Looking at Table-4, cable sample F even though has fairly high copper content, high wire guage, has low and non-varying number of twists per meter at all. C was poor in copper content, wire guage and number of twists per meter. Sample cables A, E and D even though have varying number of twists rate but the twists rate are low. G, H, I and B have high and varying number of twists rate.

4.2 Wire Guage (size)

The larger the size of the center conductor, the higher conductivity it will have. Why? Because the larger the gauge, the lower the resistance. The lower the resistance, the less signal loss a cable will have. The less signal loss a cable will have, the more signal that makes it to the display from the source [1]. This was corroborated by Kish[14] that, "The reason for the larger conductor size (approx. 23 AWG) is to provide a lower insertion loss (also called Attenuation) over the specified frequency range. A lower Insertion Loss means a stronger signal at the receiver compared with Category 5/5e". This improves the noise immunity to external and internal noise sources, Paul reiterated.

Table-3: Micrometer Screw Guage readings of the sample cable wires

Cable Name	Wire-1	Wire-2	Wire-3	Wire-4	Wire-5	Wire-6	Wire-7	Wire-8	Average Wire Guage
A	0.5	0.5	0.52	0.52	0.51	0.5	0.53	0.52	0.51
B	0.53	0.53	0.54	0.55	0.53	0.53	0.54	0.54	0.54
C	0.51	0.52	0.52	0.5	0.52	0.5	0.51	0.51	0.51
D	0.54	0.53	0.51	0.51	0.51	0.5	0.51	0.51	0.52
E	0.53	0.53	0.53	0.53	0.53	0.53	0.52	0.53	0.53
F	0.56	0.56	0.56	0.56	0.54	0.54	0.54	0.54	0.55
G	0.53	0.53	0.53	0.52	0.54	0.54	0.54	0.54	0.53
H	0.53	0.54	0.54	0.54	0.53	0.53	0.54	0.54	0.54
H	0.52	0.53	0.54	0.54	0.53	0.54	0.54	0.53	0.53

The Micrometer Screw Guage readings (diameter) of the sample cable wires were taken and recorded above. The sample cables with low copper content still have low average wire guage (A, C, D) of 0.51, 0.51 and 0.52 respectively with the exception of H that has a high average wire guage of 0.54. But then, the others with high copper content still have high average wire guage (B, E, F, G and I). This, to some large extent, corroborates the fact that those cables with poor copper content are still poor in size (diameter) while those that have high copper content continued to maintain their high quality in wire guage (size).

4.3 Ease of Crimp

Once the outer rubber jacket is stripped, wires' edges being trimmed and arranged for crimping, the organized and straightened wires begin to sag and cannot stand straight on their own (which could be due to quality of copper cable). At insertion of the eight wires into the RJ45 connector, they do not easily fall into the eight appropriate slots meant for them in the connector so some wires pass through and some do not. They are hard to crimp.

4.4 Conveyance of Power

Sometimes one needs to install outdoor router(s) high up on the roof or somewhere out of the reach of the available mains. One might need to carry power for router(s) with the same cable carrying data which should not be more than a distance of ten (10) meters for optimal result. Many of the current UTP rated category 6 cables cannot be used even though category 5 & 5e were used in the earlier times. A case of a project we installed for a faculty in my Institution convinced us. We initially thought the new router used was defective from the supplier but we later discovered that the UTP cable, rated category 6 being used, was defective and refused to carry power for the router.

4.5 Defective Manufacture processes

Even before being inserted into the connectors, some UTP cable wires' coded insulation begin to peel from the wires, leaving the wires naked. This, should not have being. This could be due to defective manufacture processes.

5. CONCLUSION

It is important that we realize that Unshielded Twisted Pair Cable (UTP) has unalloyed values in the networking world. It is affordable, slim (does not fill up trunks and pipes quickly) and now can push and scale evenly with 10 Gigabyte data which made it convenient for its use in data, audio and video applications.

We had taken nine (9) samples of UTP cable in the Nigerian UTP-cable market and have conducted elemental analysis with the use of Minipal 4 spectrometer, wire guage (diameter) measurement using micrometer screw guage and twists per meter count (per pair of cable). The main and basic constituent element of a UTP cable is copper, the bigger the wire guage the lower the resistance and the lower the resistance the less the signal loss. The less the signal loss the better the performance of category six cable. Also, the UTP four (4) wire pairs are not supposed to be same length and so the number of twists per meter should be different and high such that crosstalk among wire pairs can be reduced to the barest minimum. Three of the samples (A, C, and D) fall short of the three tests out-rightly (low copper content, least wire size and low & non varying wire pairs twists rate). Another sample (H) has very low copper content (88% Aluminium) but has very good wire size, high and varying twists rate per meter (copper should be basic element of UTP cable). Even though samples A and F have fairly average copper content (66% and 51% respectively) and good wire size, their wire pair number of twists per meter are low and non-varying (somehow will be susceptible to crosstalk among wire pairs). Only three of the UTP cable samples (B, G and I) passed the three tests having high copper content (85.0%, 73.1% and 70.1% respectively), high wire guage and high & varying twists per meter. This pass percentage of the tested samples is approximately 33.3%, which is low. This then implies that most UTP cables shipped into Nigeria are not adequate in quality.

REFERENCES

[1]. Comprehensive Cable 2012) Cable Construction Basics: Under the Jacket and Beyond http://webcache.googleusercontent.com/search?q=cache:lruQbN9N_0sJ:compcontent.com/cms/cable-construction-101-understanding-the-basic-components-of-common-cables-2/+&cd=9&hl=en&ct=clnk&gl=ng

[2] Chris Cavette (2007) How Copper is made-material used, processing steps, products, Raw Materials, The Manufacturing Process of Copper, Quality Control, Byproducts/Wastes, The Future www.madehow.com/Volume-4/Copper.html#ixzzisofY6FdV

[3]. Chris Shore (2014) What is UTP Cable www.ehow.com/about_5207438_utp-cable_.html#ixzz1klDNFlye

[4]. Cisco Systems, Inc (2003) CCNA: Network Media Types http://www.ciscopress.com/articles/article.asp?p=31276

[5]. Cisco Systems, Inc. (2004) Carrier-to-Noise Ratio in Cable Networks White Paper http://www.cisco.com/application/pdf/en/us/guest/products/ps2209/c1244/cdccont_0900aecd800fc94c.pdf

[6]. Diaverty (2009) Network Cable Testing- What Causes Data Loss? OPENXTRA Limited, Building Society Chambers, Wesley Street, Otley, LS21 1AZ http://www.openxtra.co.uk/articles/network-cable-testing-causes-data-loss#ixzz1ySEqLJNV

[7]. Elliot Davis (2012) ECOSSE REFERENCE Cable Info http://www.ecossecables.co.uk/Tables/faqmainframe.html

[8]. Catherine Bowers (1998) Ethernet Cable Length & Computer Performance www.ehow/facts_7548706_ethernet-cable-length-computer-performance.html#ixzz1mY4xwuJ6

[9]. Extron Electronics (2006): Optimizing Performance Extron Electronics Sending A/V Signals Over Twisted Pair Cables Headquarters in Anaheim, CA. http://www.extron.com/download/files/whitepaper/tp_opt_wp.pdf

[10]. Galen Udell (2006) Discussion Paper: Category 6 versus Category 5e Cabling Systems and Implications for Voice Over IP Networks Belden CDT Networking http://www.newark.com/pdfs/techarticles/belden/DifferenceBetweenCat6Cat5Standards.pdf

[11]. Hitachi Cable Manchester (2009) The Facts About Category 6 and 6A Cabling Hitachi Cable Manchester, 900 Holt Avenue, Manchester, NH 03109 White Paper, January 2009. http://www.accu-tech.com/Portals/54495/docs/hcm%20copper%20cable%20selection%20guide%202010.pdf

[12]. Jay Potter (2002) Ethernet Over Standard Category 5 Cable: Functional Testing &eE Certification International Academy of Science https://www.wband.com/resources/whitepapers/Ethernet_Without_Rewiring_White%20Paper_802.pdf

[13]. J. Griffith et al (2000) Cable Parameters and Acoustic Probe Performance E & H Resources, Inc., Phoenix, AZ 85048; Weidlinger Associates, inc., Los Altos, CA 94022 and NUWC, Newport RI, 02841. Preprint of an article in the 2000 IEEE Ultrasonics Symposium Proceedings

[14]. Paul Kish (2002) Category 6 Cabling Questions and Answers NORDX/CDT Cable Designs Technologies http://www.belden.com/docs/upload/what_is_category_6_q-a.pdf

[15]. Mike LaPlaca (2012), "Tech Brief- Effect of Return Loss& Impedance Stability on Premise Wiring Systems at Extended Frequencies"

Vol 7. No. 3 - September, 2014
African Journal of Computing & ICT

www.ajocict.net

A Social Engineering Detection Model for the Mobile Smartphone Clients

A.A. Ojugo
Dept. of Mathematics/Computer
Federal University of Petroleum Resources
Effurun, Nigeria
ojugo.arnold@fupre.edu.ng

A.O. Eboka
Department of Computer Education
Federal College of Education (Technical)
Asaba, Delta State, Nigeria
andre_y2k@yahoo.com

ABSTRACT

The continued advances in both Internet and mobile smartphones has up-scaled client adoption of Internet-based data processing activities like e-banking etc. These, have also necessitated growth cum deployment of mobile applications across platforms to help clients accomplish data processing tasks by harnessing its many benefits. Yet, it has also exposed clients whom are today, constantly besieged by data security threats that are not limited to social engineering attacks. Thus, towards data residence, sensitive and proprietary – clients have become bothered with the exposure of their smartphones to possible data loss, theft, manipulation and inspection etc – if they are to further adopt these new paradigms. These, leaves fears and doubt in the mind of clients (as data owner) and reduces their trust-level in such services. We propose a client-trusted security model for smartphones employed in banks that aims to increase clients' trust-level in the adoption of mobile banking – making it more dependable as the framework seeks to address security threats with transaction authenticity and message authorization. Results shows framework is capable of increasing client's trust level in relation to social engineering attacks with about 72% as implemented over their firewall by the banks (for Internet connectivity as well as ported on a community-cloud) for user access.

Keywords—phishing, whaling, framework, social engineering, financial institutions, fraud, vishing, smishing

African Journal of Computing & ICT Reference Format:
A.A. Ojugo & A.O. Eboka (2014). A Social Engineering Detection Model for the Mobile Smartphone Clients.
Afr J. of Comp & ICTs. Vol 7, No. 3. Pp 91-100.

1. INTRODUCTION

SIgnificant advancement with the field of information and communication technology (ICTs) has since climaxed with exponential growth and advancements of the Internet that continues to beam its benefits to its plethora of clients or users. These can today, be seen to have permeated into the fabrics of our daily lives and activities, as we employ ICTs for personal, business and recreational purposes. ICT continues to advance efficient dissemination of data for effective decision making. A corresponding sine-qua-non effect is the myriad of threats that seeks to exploit the inherent vulnerabilities in these advances of ICT and associated technologies for many of its naïve users. These challenges and threats manifests in various forms or ways presenting itself as misleading items of benefits to many 'unsuspecting' users, and aimed at defrauding them [1]. ***Fraud*** is a criminal act, perpetrated via embezzlement, larceny and theft in which a criminal employs falsehood to benefit from an unsuspecting patron, or from assuring victim of great returns (if it is aimed at a financial transaction) such that the victim relies on such falsehood.

A transaction is exchange of goods and services for gains or money deliverables [2]. Criminals find it easier to exploit users of a service than exploiting a web application or network connection, and many organizations who invest in highly sophisticated security often fail to adequately address their biggest vulnerability, which is deception of their employees. However, criminals increasingly use deceptive means to exploit corporate business practices and circumvent controls so as to trick unsuspecting clients into sending money or diverting payments to imposters [2][3].

Social engineering (not a new paradigm) has steadily grown with no-end-in-sight. Its continued growth borders on human nature of trust instincts, on which hackers manipulate human emotion and ultimately, exploit this trust to steal valuable information. Common technique for achieving this feat are: phishing, vishing, smishing etc – with the most popular being phishing.

It uses social engineering and technical subterfuge to defraud an online account holder of their financial information by posing as a trusted identity. Phishing can be executed via multiple means including: spoofed emails and phone calls, web link manipulation and forgeries, man-in-the-middle chat, covert redirect etc – all aimed at convincing a user to divulge confidential data and/or participate unknowingly in fraudulent transactions [4].

These attacks are mostly targeted at clients that employ Internet daily, from which a greater multitude continues to adopt the mobile smartphone. These attacks have thus risen in number for smartphone clients – resulting from the increased growth of user access to mobile smartphones from 42.5% in 2013 to 78.9% by 2013 and the advent of Androids is made smartphones a preferred choice over personal computers due to its design, portability, speed, functionality and ease of Internet access. All these, continually pose significant threats with its high vulnerability rate and security risk to user data. Consequently, these have its range of implication to work-related functions and business issues as it often exposes sensitive data to adversaries. Even with advancement cloud computing to the rescue, data security challenges still persists with many impediments into the full realization of its potentials and benefits as it seeks to explore storage capability to guarantee data recovery of user data at various levels [2].

Data and applications shared both on the ***Internet*** and as extended via ***cloud*** computing – really resides in systems that the clients have limited or no control over. And though, in both instance, transaction or message authentication is done by client – there are often no checks or measures that allows for authenticity feedbacks or handshaking is performed between the client and service provide, to confirm from each other the validity of a transaction or message. This has become a major challenge, and today, is quite responsible for most security issues associated with both cases of the Internet and as extended via cloud technologies. Other security concerns for service providers include: (a) possible harm to an organization for public distributed access, (b) recovery cost implication from such harm, and (c) other associated risks that may result in service failure or denial of service. These concerns (should) raise questions in the minds of its many prospective clients – since the inability to resolves these challenges, leaves them bewildered and insecure towards adopting these services, irrespective of its many benefits even in the foreseeable future [2][4].

Thus, the ***idea*** or objective of this study is to develop and deploy a framework model for early detection of fraudulent activities spawning from or geared towards social engineering attacks that are aimed at smartphones clients.

A. Social Engineering

A client (or end-user) of a service can perform a transaction by send a message to another user or executes control message as sent to such a client by the service provider. Thus, client is given access to a plethora of services by the service provider. With social engineering attacks, many see the obvious victim as the client engaged in using the service; But, actual target of the attack is the service provider or organization. For financial services companies (banks), many of her services to clients can be potentially damaged with social engineering schemes like phishing. In 2015 alone, over 67% of phishing attacks were trading on names of banks and financial organizations, while 61% of the phishing attempts were specifically targeted at the financial credentials. Phishing exploits are becoming more successful and even now, exceedingly more difficult to detect (as most of them go unreported). 23% of above fact were lured to open phishing messages; while, 11% were lured to open its attachments [5]. These attacks may be aimed at the unsuspecting user on the obvious; But, the actual claim is that it continually undermines the client's and/or user's confidence in a brand, puts the client at great risk of identity theft (to mention few) and consequently, incurs huge financial losses, debts and deficits to the financial sector or industry. As phishing and other social engineering attacks escalate, financial institutions are soon enmeshed in recruiting new means to track and mitigate these attacks and its potential damages; While, they constantly battle to remain updated in both the finance and ICT world [2].

B. The Android Smartphone Framework / Platforms

Smartphones are today, a preferred device of choice over the personal computers for some data processing activities due to its ease of Internet access, portability, durability and speed of processing (with its seemingly user-satisfaction of upload and download time, and to mention a few). Many of these employ various operating systems that allow it to perform many of the much user-needed feats of data communication and processing activities. Examples of such operating systems today include: Android, Windows, Symbian, iOS, Blackberry etc.

[6] Android is today, a leading platform for mobile devices owing from its open source feat to distinguish it from other platforms such as Blackberry, Windows Phone and iOS. It is not a specification nor distribution of traditional Linux, neither is it a collection of replaceable components or chunk of software ported on a device. Its open source platform is built by Google with OS, middleware, and applications for mobile platforms based on Linux kernel – enabling developers to write apps majorly in Java with support for C/C++. A major success is in its licensed under Apache2, allowing third party porting-developments to the platform. Since its release, it has been constantly improved either in feats, supported hardware, and at same time extended to new device types from the originally intended mobile ones [7]. Recent efforts are geared towards enhancing for real-time capabilities, to be employed in variety of other embedded systems [6][8].

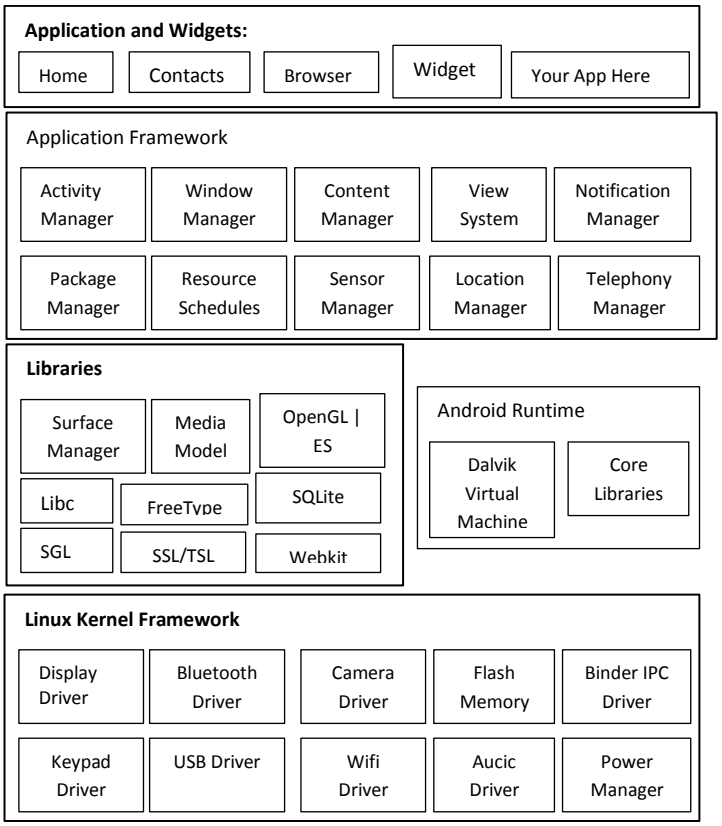


Fig 1: Android OS Platform

[9][10] in "Google Android and Linaro Android SDK" note that the Android Platform is an eco-system layer of software component implemented on the mobile smartphone hardware as thus (see fig 1) and at whose core are these features:

a. Linux OS provides basic functionality such as security, process/memory management and networking to support vast device drivers. It handles human machine interfaces, file systems, network access etc. Its kernel is modified by Google to use low memory killer, specific inter-process communication system, kernel log feats, shared memory system and many other changes as developed. It runs on standard Vanilla Linux, merging specific changes into its kernel. Recent release aimed at real-time Linux kernel is v4.0.3 (Ice Cream Sandwich).

b. Library with Google's *libc* called *Bionic*, media/graphics (OpenGL|ES), browser-webkit and light-database SQLite. DVM (Dalvik Virtual Machine) completely differs from Sun's JVM and uses register based byte code to conserves memory, max performance and can instantiate many of its apps multiple times, with each app having its own private copy running. DVM uses Linux for memory management and multi-threading to support the Java language.

DVM uses *bionic* (not compatible with *glibc*) so that its native libraries are faster to implement with small custom *pthread* to support services such as system and logging capabilities. Writable data segments are small so as to be loaded into memory with each process. This keeps code size small so that Linux loads only once, all read-only pages. *Bionic* is used: (a) to avoid inclusion of GPL code at user space level in its platform where BSD is used, and (b) for small memory footprint devices with high speed CPUs at relatively low frequencies. *Bionic libc* does not handle C++ exceptions (though omitting such lower level exceptions pose no problem as Java is Android's primary language. It handles exceptions internally). *Bionic* has no priority inheritance for mutexex as implemented in *glibc*. Available in its kernel and accessed via own library in system calls, its lack of priority inversion disqualifies it for real-time capability as applied in robotics/automotive. Google's reason for a complete new VM from scratch as accomplished with DVM's register-based byte code is to reduce patent infringement risk. Thus, existing real-time apps modified for JVM cannot easily be ported to DVM.

c. Application Framework provides higher-level services to apps such as Java classes amongst others. Its use can vary between/with varying implementation.
d. Application/Widget are Android routine distributed apps such as email, SMS, calendar, contacts and Web browser.

C. Security Concerns and Controls

Many of the security concerns and risks associated with both the Internet and extended by cloud technologies can be safely handled by organizations via planned risk management in biz processes and activities. Examples of these include: (a) the right to choice of service provider, (b) the legal responsibility that must be accepted by service provided, the threat of access to intellectual properties, and (c) content of disaster recovery documentation [11-12]. But, lack of control on the physical infrastructure is responsible for most of the security issues that arise in both instances of the Internet and cloud. Furthermore, clients are ignorant of the physical location of their stored data in the distributed environ as well as what security mechanisms are in place to guard their data and defend them against hackers and other adversaries [13].

Also, security issues relate to web services and web browsers for mobile smartphone clients on the Android platform, is addressed in [6]. These can be adapted to other platform technologies as the most common attacks on web services is XML Signature Element Wrapping with XML signature used in authentication of a transaction or message [12].

Security controls are required in all ICT-enabled environ. In addition, the Internet presents different risks to different client as they access it due to the services requested for, operations to be performed as well as technologies of smartphone devices that are associated with it. Internet security control models can be applied: (a) to applications via firewalls, (b) to data via database activity monitoring, (c) managing infrastructure via configuration management and monitoring, (d) to intranets and other forms of network via firewalls, and (e) data storage via encryption schemes. Use of traditional security controls like access controls and encryption, monitoring of large internal data migrations with database/file activity monitoring, and monitoring of data movements via the Internet with URL filters and data loss prevention [6].

2. MATERIALS AND METHODS

Dataset Used

Available social engineering data is as in Table 1 and table 2 – represented as in Fig. 1 and fig. 2 respectively.

Table 1: Target List of Social Engineering represented in Fig 1

No	Types of Organisations	Percentage
1	Social Sites	15.2
2	Financial and Banking Services	34.4
3	Portals	19.7
4	Military	9.8
5	Government Officials & Top Management Personnel in Government Parastatals	8.3
6	Others	12.6

Table 2: Kaspersky Target List of Phishing Attempts

No	Phishing Attempts On	Percentage
1	Financial Credentials	21
2	Recipients Opening Phishing Messages	23
3	Recipients Opening Phishing Attachments	11
4	Vishing, Smishing, Whalling and Others	45

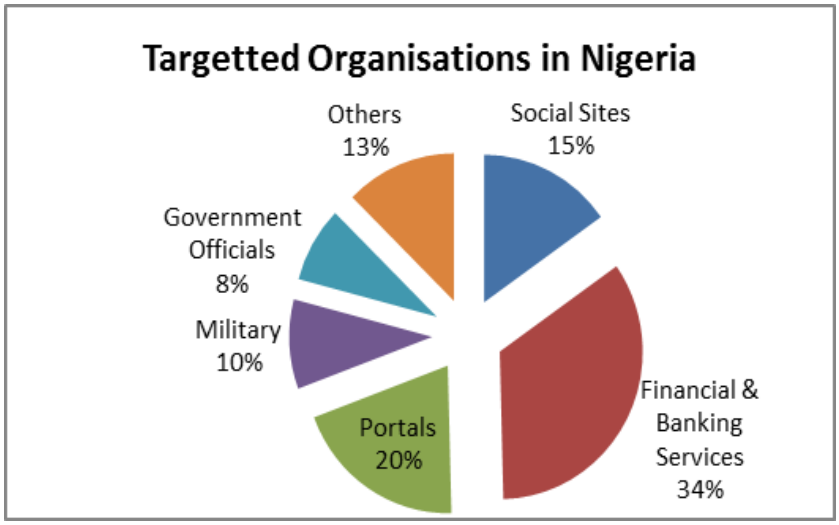


Fig. 1: List of targeted organs by Social Engineering Attacks

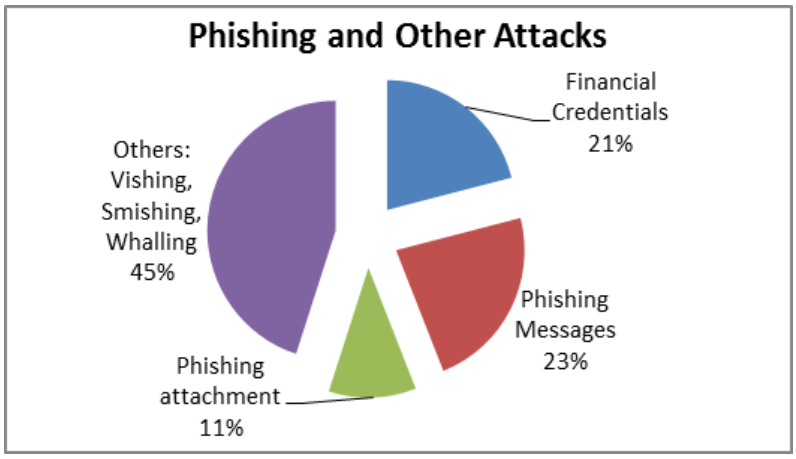


Fig. 2: Phishing Attacks and Others

The ***study*** seeks to model authentication and authorization framework for smartphone clients in financial transaction. This will aid clients to evade and avoid socially-engineered attacks and fraud aimed at smartphones.

D. Statement of Problem

The problem statements are as follows:

1. Advancement in Data Security continually beckons on the exchange of data to update awareness of techniques used by hackers as well as exchange of data in fraud detection, which is often limited as it is unwise to describe in public domain (fraud detection techniques in great detail). As a dual effect, it will further equip organisations, users and hackers with adequate information required to combat as well as evade significant detection (for hackers). Thus, we employ statistical fraud detection method and heuristics as in Section III.
2. A major challenge clients face in accessing the Internet, is that of message authenticity – as it is difficult for a user to differentiate between a service provider's message (as is case of 'targeted' financial institution's websites) and a fraudulent one, leaving them exposed to fraud and attacks.
3. Unavailability of social engineering (fraudulent) datasets due to unreported cases as well as its ***uncensored*** results, and the non-establishment of regulatory organs to monitor such activities, makes fraud detection, its techniques and consequent studies, difficult to assess. Also, the available dataset are plagued by noise, ambiguities and impartial truth. These, must be resolved

via accurate classification models and algorithms that seeks to efficiently classify observations and expected values of rules generated from the dataset. This is resolved in Section III/IV.

4. Another major challenge is that the banks have no control on authorization decisions as a hacker can easily convince a client to generate a signature and authorize a fraudulent transaction without the bank's involvement. This is because in a typical process, the bank has zero control over the authorization decision – even with the use of *Digipasses* (hardware mini tokens). The client completely authorizes any of his banking requests at any given time – including fraudulent ones. This is a major reason why phishing continues to be successful.

The goal and objective of the study is to provide a framework that achieves the following:

a. In using a mobile banking platform – the client is required to configure the device used and synchronize data with the banking 'mobile-device' database, which registers and authenticates a client's banking details with the mobile device used. This aims to remove the issue of trust and trust decision, out of the hands of the client and ensuring that when a client request is made to initiate a transaction, the bank in tandem with the client initiates and grants access to a transaction signature request. This can be achieved using a cryptogram.
b. The bank establishes a secure communication channel between the client's device and their platform that enables the message and transaction to be authenticated.
c. Ensure that the bank controls the authentication process as the client initiates any and all transactions.

3. PROPOSED EXPERIMENTAL MODEL

Client data stored in smartphones have become easy targets for malicious attempts. Clients may not understand security features as provided in varying forms, needed to be in place as they constantly access Internet and other computing platforms. Thus, study proposes a reliable modeled framework that seeks to help users uncover and detect socially engineered attacks that are aimed at smartphones.

The framework seeks to bring a sense of control to many of the Service providers (infrastructure and services) as well as grant a failsafe to mobile smartphone via a single platform via mobile computing. It is an extension of the miniaturization process and faster computing on Moore's law, bringing about dependable and secure data storage capability to users via portable mobile devices.

The framework is designed to thwart efforts geared towards social engineering attacks via phishing, vishing, smishing and whaling (to mention a few). It is now customary that very often, most clients use ***digipass*** (or ***tokens***) to authenticate their every mobile banking transactions or messages (between them and their banking institutions) from their trusted device (smartphone) via either the hardware or software form for message and transaction authentication as well as client authorization. As thus:

a. Cryptogram – [6] GSM and CDM/FDM are grouped under second-generation systems. *GSMs* are developed to resolve incompatibilities in the cellular technologies so that a number of subscriber units can be used. Its basic feats are: subscriber Identity Module (SIM) portable smartcard-like plug-in device that stores a user's identification number, the network a client is authorized to use, encryption keys, and other info specific to a user or subscriber. Terminals are generic until SIM is inserted so that the subscriber unit and SIM roams; Transmissions between base transceiver and user is encrypted with ***A5*** cipher; while, ***A3*** cipher helps authenticate calls making it private to a user. GSMs support data and image services based on ISDN (Integrated Service Digital Network) model with user data rates upto 9.6kbps.

   The advent of third-generation communication system allowed mobile subscribers and personal communication services to incorporate set of standards and services unto the mobile unit (making them smartphones) so that it can then support data in form of voice, files, image and video. Though, housed within the device is the cipher (A5 and A3) that helps to authenticate data sent between devices and the service provider. However, banks can provision for extra layer of cryptogram that allows further ciphering so that the data packets (messages or transaction) initiated by a client as sent via a client-trusted device encoded for adequate authentication by the bank. The cryptogram contains all transaction data, including: registered device in use, transaction amount, and recipient account details.
b. Digipass (Hard/Soft tokens) addresses clients' transaction authorization and message authentication weakness via and by assuring the bank that only a registered/authorized client can authorize a legitimate transaction, so long the bank is able to verify the user details which includes (not limited to): registered device in use, account details and benefactors (recipient from another bank) etc [14-16].
c. Data Configurable Model allows for trusted computing via virtualization. Virtualization simply is the process of decoupling hardware from an OS on a physical machine. The Internet (and Financial institutions today), via mobile banking provide users with multiple isolated environ or sites, known as virtual machines (VMs) on a single host. VM is a virtualized representation of a physical machine that runs or is maintained on a host machine by software virtual machine monitor (hypervisor), to provide trusted computing, a mechanism that allows financial institutions to verify their security posture through hardware/software controls. A key component is the trusted platform module (TPM), which is a cryptographic component that provides a root of trust for building such a trusted computing base. Its goal is to move cryptographic computations into a locked virtual area, which is not under control of entities on the host platform. TPM works only in non-virtualized environs. A virtual Trusted Platform Module is provided according to standard specs by creating an instance of TPM for each VM on a trusted

platform. All these layers of security – helps to act as measures to ward-off hackers and intruders [17-20].

d. Web Services API – Framework seeks to implement an AES-256 encryption, which is easily ported and supported by SSL/TSL. We adopt this for: (a) it is computationally secure against brute-force, (b) flexible, (c) small-size Java codes allows support for C-language, (d) memory size required is small as ported on Android platform as it has no effect on the performance of memory and speed of the smartphone, and (e) ease of integration as implemented with Java and support for C-language into its web browsers with ease of connectivity. Also, the Chrome and Firefox Web-browsers can be used as they all allow AES-256 encryption. Since, data transfer over the Internet between a client and his/her service provider remains unprotected, no matter how good SSL in use is.

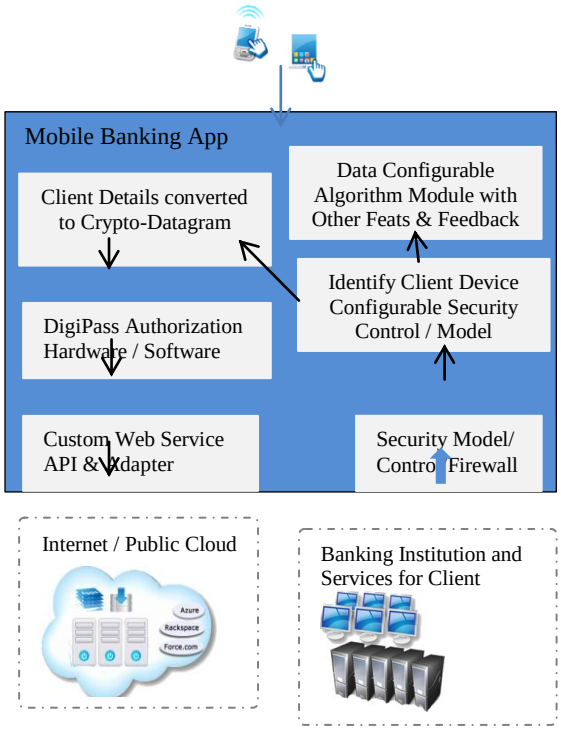


Fig 4: Application and Data Model For Smartphone

The study provides a framework security model for clients using smartphones, and operable on the Android v2.4 platform with support for web-service to allow easy access to Internet connectivity and connection to remote server/cloud-services via API call (adapter). Framework masks all technical nuances between application-model and data-models such as session management, connectivity-issues, transaction authentication and message authorization between a client and his/her bank. Its security is first handled via the digipass that generates a user personal identification number (PIN) from the hardware device (called a token). It is worthy to note that the device is already registered to the user so that whatever token generated can allow the user to log unto his account via the A5 and A3 ciphers provided natively by the mobile smartphone device [21-24]. Furthermore, the client's smartphone device is then used to log unto the bank's platform bypassing their firewall – having been granted access via generated token so that sent message are authenticated and initiated transaction are verified by the bank to originate from an authorized client. The system then uses the cryptogram to ensure that transactions and messages are properly formatted according to the bank so that they are allowed to efficiently be transmitted via the AES-256 crypto-system on SSL/TSL as applied to all data. Its client end-to-end encryption solution uses AES-256 to protect its data; while SSL protects username and password (see fig. 1).

Tools used for the development of this native app include Android SDK, Apache XAMPP and Google's Android Studio. This native app is enabled and ported on any Android platform from v2.2 with forward compatibility.

E. Experimental Implementation and Findings

The proposed framework was tested through comparison and evaluation with live and running scenario of a community-cloud provider (as well as connected over the Internet) and a user. The user is natively connected to the Internet via his/her smartphone device, and also subscribed for IaaS services via

FUPRE community-cloud to US-based cloud service provider for purpose of installing some proprietary mobile smartphone applications. Product details for the cloud provider, options available and values set are as in Table 3; while, the provider's level of Client Trusted Process Model of proposed model requires documentation (as in Table 4).

Table 3: Smartphone Internet Provider's Product Options

Product Details	**Options set**
Registration Date:	5/27/2012
Product/Service:	Online Traders - VPS Value Edition
CNS Subscription ID:	118223
VM:	VM118223.tradersvps.net
IPv4 Address:	173.228.134.65
Number of Snapshots:	1
CPU Cores:	2
RAM (MB):	640
DISK (GB):	20
Two-factor Authentication:	No
VNC:	Do not install VNC
Operating System OS:	Traders VPS Windows 2003 (x86) Enterprise Edition R2
Language:	English
Datacenter:	NYC
Payment Method:	MasterCard, Visa & Am- Express
First Payment Amount:	$30.00 USD
Recurring Amount:	$30.00 USD
Next Due Date:	8/27/2016
Billing Cycle:	Monthly
Status:	Active

The location of the datacenter, the description of the Virtual Machine, monthly rate paid by the user, the security control provided by the provider and other product information are shown in Table 1. It is seen from Table 1 that the only security control provided by provider is "Two-factor Authentication".

Table 2: Security Control with Documentation Proposed Connectivity

Security Control Name	Documentation Label and Documentation
Two-factor Authentication	**D1**: As a client initiates a log in from registered smartphone device unto the bank's platform, it requests that the client be verified at each log-in request. Thus, the bank verifies if the client is an authorized user via the token generated by the digipass and authentication message. Also, each client is prompted to re-verify at each log in session and/or having successfully logged off from the banking platform. The user can also request a token be sent to him/her for verification via SMS if such a user does not have a the hardware token (though this would not be the usual case). This currently exists for most cloud platforms
Use secure provisioning and Secure Migration Protocols	**D2**: These protocols prevent data from ever being sent to malicious hypervisor, virtual machines and host whenever a new virtual server is requested on the cloud. It puts a verification mechanism in place to ensure that that attacks against the virtual environment of your stored application or data will not be performed by an unapproved Operating System.
Virtual Trusted Platform Modules	**D3**: Virtual TPM protects its internal data from being accessed by the host environment, hypervisor, and all other virtual environments on the platform and puts a protection in place to prevent itself from being cloned and it is maintained in a secure location under your full physical control.

Security control with documentation for proposed framework using simulation is shown in Table 4. Proposed framework uses 3-security control systems at point of data configurable algorithms. The user trust level increases with the number of operational and well document security controls. Comparing the initial values with Table 4, the proposed framework is capable of increasing the trust level of the client by about 72 % when compared to the existing cloud system.

The users trust can increase above this value as more appropriate security controls are put in place to clear the user's doubt and enhance the trust level.

F. Benefits Accrued to Client and Institutions

Some benefits to be provided by the framework include :

a. Help combat social engineering and other online banking threats
b. Mitigate human risk in online banking transactions
c. Works with push notifications to immediately alert a client when updates and other information are made available by the service provider
d. Enables flexible deployment on any screen as the client can work between his trusted device and computer system – and vice versa.
e. Improves user experience of mobile (smartphone) online banking via its scan and sign that enables fast adoption and brings both safety and simplicity to user when signing transactions.

G. Related Study

[6] uses integrated framework for dependable community-cloud computing for smartphones. Study provides a support tool ***PushCloud*** that allows clients access a user account with the capability to sign-in and perform backup functions on contacts, messages, picture files, documents, videos and recorded voice amongst others via their smartphone unto the cloud with access granted via their service provider. The system proffers benefits such as the ability to pool together cloud service providers, provides a user with a cross-platform with minimal price difference as well as aim to address security-related issue from a user's end via AES-256 encryption on the integrated cloud model. This it does while exploring storage capability to guarantee recovery of data from a remote server (BDC) for back-end as well as front-end data storage ease [25-28].

4. CONCLUSION AND RECOMMENDATIONS

In this study, we model a client smartphone framework for detecting and warding off social engineering attacks and fraud. The model is designed to help address and thwart social engineering attacks by taking the 'trust' decisions out-of-the hands of the client, ensuring that while only a client initiates a transaction, authentication and authorization is domiciled with the bank to verify that the said transaction is by an authorized client via his trusted device (smartphone) [29-31]. Although deception fraud attacks will likely continue with increased frequency and sophistication, most organisations will equip themselves with the capability to minimize and mitigate all risks they and their (prospective) clients will encounter. While, the good news is that it does not require a massive ICT budget; It does however require more of commitment on the part of the financial institutions (banks) to invest time and resources into employee education and training (as possible threads for threats is most likely to come from them as they are better disposed to the flaws inherent in their system) as well as invest into their system to safeguard their future with their clients. As a result, there is no excuse not to implement basic deception and/or fraud security measures, and every company (financial organisations) should consider insurance as an important component of their overall risk management program[32-35].

REFERENCES

[1] Yeboah-Boateng, E.O and Amanor, P.M., *Phishing, smishing and vishing: an assessment of threats against Mobile Devices*, Journal of Emerging Trends in Computing and Information Sciences, 2014, Vol. 5, No. 4, pp297-307

[2] Ojugo, A.A., D. Allenotor., D.A. Oyemade., O.B. Longe., C.N. Anujeonye., *Comparative study for credit-card fraud detection models*, African Journal of Computing and ICT, 2015, Vol. 8, No. 1, Issue 2, 2015, Pp 15 - 24

[3] Symantec Corporation, *Internet Security Threat Report*, 2015, Vol 20, www.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-reportvolume-20-2015-social_v2.pdf

[4] Alnajim, A.M. *Fighting internet fraud: anti-phishing effectiveness for phishing websites detection*, unpublished doctoral dissertation, 2009, Durham University, Durham

[5] VASCO Data Security: *Social engineering vulnerabilities Report*, 2015, Vol 12, www.vascodata.com/mktginfo/whitepaper/ISTR/internet-security-threat-reportvolume-20-2015-social_engineering_phishing..pdf

[6] Ojugo, A.A., Aghware, F.O., Yoro, R.E., Yerokun, M.O., Eboka, A.O and Anujeonye, C.N., *Dependable Community-Cloud Framework for Smartphones*, American Journal of Networks and Communications. 2015, Vol. 4, No. 4, pp. 95-103. doi: 10.11648/j.ajnc.20150404.13

[7] Maia, C., Nogueira, L and Pinho, L.M., (2010). Evaluating Android OS for Embedded Real-Time Systems, Proceedings of 6th International Workshop on Operating Systems Platforms for Embedded Real-Time Applications, Brussels, Belgium.

[8] Tapas Kumar, K and Kolin, P., (2010). Android on Mobile Devices: An Energy Perspective, IEEE 10th International Conference on Computer and Information Technology, Kuala-Lumpur, Malaysia.

[9] Pernel, L, Fayyad-Kazan, H and Timmerman, M., (2013). Android and real time application: take care, J. Emerging Trends in Computing and Info. Systems, 4, Special Issue ICCSII, ISSN 2079-8407.

[10] Agam, S., (2011). Google's Android 4.0 ported to x86 processors, retrieved online via: http://www.computerworld.com/s/article/9222323/Google_s_Android_4.0_ported_to_x86_processors.

[11] ISACA White Paper, *Cloud Computing: Business Benefits With Security, Governance and Assurance Perspectives*, 2009, p1-10 [online]: www.klcconsulting.net/security_resources/cloud/Cloud_Computing_Security_&_Governance-ISACA.pdf Accessed July 2016.

[12] Babu, G. N. K. S., Srivatsa, S. K., "Security And Privacy Issues in Cloud Computing", International Journal of Engineering, Business and Enterprise Applications, 2014, pp 145-149.

[13] Jensen, M., Schwenk, J., Gruschka, N and Iacono, L. "On Technical Issues in Cloud Computing". In IEEE International Conference, Honolulu: Piscataway, 2009.

[14] Banday, M.T., Qadri, J.A. .*Phishing - A Growing Threat to E-Commerce*, The Business Review, 2007, ISSN: 0972-8384, Vol. 12, No. 2, pp. 76-83.

[15] Bankash, A., *For detailed information on PWSteal*. 2015, http://securityresponse.symantec.com/avcenter/venc/data/pwsteal.bankash.a.html.

[16] Barbieri, R., Bruschi, D and Rosti, E., *Voice over IPsec: analysis*,2012, [online]: available at www.securityresponsekaspersky.com/pwsteal.roshi.html

[17] Brar, T.P.S., Sharma, D and Khurmi, S.S., *Vulnerabilities in e-banking: A study of various security aspects in e-banking*, International Journal of the Internet of things, 2016, Vol. 5, No 3, pp 121 – 134.

[18] Chaturvedi, A and Meena, A., *Analysing the impacts of phishing and vishing attacks in Internet banking*, International Journal of Advanced Research in Computer Science and Software Engineering, 2016, Vol. 6, No. 3, pp 16 – 21.

[19] David, L., *Phishing expedition at heart of AT&T hacking*, San Francisco Chronicle, 2006.

[20] Davis, M., "*280 Kansas City employees fall for fake hack"*, 2015, [online]: available at www.kansascity.com/news/business/technology/article16376894.html, last retrieved July 2016

[21] Corrin, A., *Spear-phishing tactics becoming more sophisticated*, 2014, [online]: available at http://archive.federaltimes.com/article/20141024/CYBER/310240013/-Spear-phishing-tacticsbecoming-more-sophisticated

[22] Federal Bureau of Investigation., *Spear phishers*, 2009, [online] www.fbi.gov/news/stories/2009/april/spearphishing_040109.

[23] Ferguson, D., *Phishing warning Beware e-mails asking for personal info*, as cited in Yeboah-Boateng, E.O and Amanor, P.M., *Phishing, smishing and vishing: an assessment of threats against Mobile Devices*, Journal of Emerging Trends in Computing and Information Sciences, 2014, Vol. 5, No. 4, pp297-307

[24] Hong, J., *Why have there been so many security breaches recently*? 2015, http://cacm.acm.org/blogs/blog-cacm/107800-why-have-there-been-so-many-%security-breaches-recently/fulltext

[25] Johnson, M., *New approach to Internet Banking*, September 2008 Technical report (UCAM-CL-TR-731), Cambridge University, ISSN 1476-2986

[26] Korolov, M., *Omaha's Scoular company loses $17 million after spear phishing attack*, 2015, [online]: available at www.csoonline.com/article/2884339/malware-cybercrime/omahas-scoular-co-loses-17-million-after-spearphishing-attack.html

[27] Marko, J., *Larger preys are targets of phishing*, 2008, [online] http://www.nytimes.com/2008/04/16/echnology/16whale.html.

[28] FireEye, "*Spear Phishing Attacks – Why They are Successful and How to Stop Them: Combating the Attack of Choice for Cybercriminals*, 2012, [online]: www.crsonline.com/contents/security/reshsecurity.html.

[29] Firefox phishing protection bypass vulnerability. Securiteam, Jun 2006. http://blogs.securiteam.com/index.php/archives/467

[30] Gartner, Press Release, "*Gartner Says Worldwide data security Spending Will Grow 8-Percent in 2014 as organizations Become More Threat-Aware*", (August 22, 2014), http://www.gartner.com/newsroom/id/2828722

[31] Purkait, S., *Phishing counter measures and their effectiveness – literature review*, International Journal of Internet, 2015, Vol. 23, pp380-420.

[32] RSA, *A Year in Review*, http://www.emc.com/collateral/fraud-report/rsa-online-fraudreport-012014.pdf, 2014, last retrieved July 2016

[33] Waugh, R., "*Big Companies still fall for social engineering hacks by phone and it's not getting better*, 2013, http://www.welivesecurity.com/2013/10/31/big-companies-stillfall-for-social-engineering-hacks-by-phone-and-its-not-getting-better/

[34] Wang, X., Zhang, R., Yang, X., Jiang, X and Wijesekera, D., *Voice pharming attack and the trust of VoIP*, ACM Transaction on SecureCOMM, ISBN: 978-1-60558-241-2, 2008, Vol. 21, pp241-321.

www.ajocict.net