

# Mining and Analyzing Academic Social Networks

Tasleem Arif

Department of Information Technology  
Baba Ghulam Shah Badshah University  
Rajouri, J&K, India

---

**Abstract:** Academics establish relationships by way of various interactions like jointly authoring a research paper or report, jointly supervising a thesis, working jointly on a project, etc. Some of these relationships are ubiquitous whereas other are hard to keep track of. Of all types of possible academic and research collaborations, co-authorship is best documented. In this paper we analyze the co-authorship based academic social networks of computer science engineering departments of Indian Institutes of Technology (IITs) as evidenced from their research publications produced during 2011 and 2015. We use social network analysis metrics to study the collaboration networks in four leading IITs. From experimental results it can be concluded that IIT Delhi and IIT Kharagpur have a close knit collaboration network whereas the collaboration network of IIT Kanpur and IIT Madras is fragmented. However, the collaboration networks of all the four IITs exhibit similar network properties as expected from any other collaboration network.

**Keywords:** Social Networks, Co-authorship Networks, Social Network Analysis, IITs

---

## 1. INTRODUCTION

Publishing research work is one of the primary activities of academics. These publications are intellectual contributions in terms of ongoing or completed work of one or more people [1]. Sharing of ideas has traditionally proven to contribute immensely to existing knowledge. This is truer in case of science. Studies [2] point towards dominance of sole author publications in the beginning of the 20th century whereas, the trend got reversed with the passage of time [3] as the percentage of sole author publications took a plunge. The rise in number of co-authors of a publication is the need of modern day research as it has become more specialized; requires clear understanding of underlying concepts of diverse subjects; has to be interdisciplinary in nature, etc. [1]. It therefore becomes next to impossible for a single person to deal with all aspects of a problem and he has no choice but to seek support and work with other people in a team setup.

In an online social network environment users establish relationships by sharing status, by way of likes, or tweets and retweets. However, these relationships are casual, whereas, on the other hand, the relationships between researchers established through various academic activities are much more formalized [4]. There are a number of ways in which academics collaborate and establish academic relationships but co-authorship is the most tangible and well documented form of research collaboration [5]. By means of being co-authors academics forms social networks [4]. In order to extract and study these networks the co-authorship data has to be mined, processed and used. Digital libraries like DBLP, Microsoft Academic Search provide a rich source of co-authorship information and make it available online. In addition, institutional websites also serve as a rich source of co-authorship information on people working with that institution. Analysis of these co-authorship relationships provides a lot of information about individuals, groups and institutions.

Orgnet.com<sup>1</sup> defines social network analysis (SNA) as a tool for analyzing relationships and flows between various entities like people, groups, organizations, computers, URLs, etc. As

discussed in [5], in social networks, nodes represents entities like people, groups, organizations, etc. whereas, the edges represents the relationships or flows between the nodes. Using SNA techniques one can have both a visual and a mathematical analysis of relationships in social networks.

Study of the patterns of interaction and communication in collaborations between various actors has already attracted significant interest from scholars [6, 7, 8, 9]. Advances in data mining and recent developments in social network visualization software have facilitated the study and analysis of intensity and dynamics of these relationships in a visual or graphical manner [9]. Representation of interactions between entities in terms of nodes and edges i.e. graphs, where nodes represents entities and edges represents interactions, allows one to apply graph theory for the analysis and understanding of underlying collaborations [9]. Such a study is capable of finding and describing the interactions at micro, macro and universal level.

In this paper, we discuss the extraction of collaboration networks to study co-authorship collaborations of people of computer science engineering departments of four Indian Institutes of Technology (IITs) over a five year period. Like any other graph the collaboration network is essentially a graph represented as  $(G=V, E)$ , where the vertices ( $V$ ) represent authors and the edges in the edge set ( $E$ ) represents the co-authorship link between these authors. Each edge ( $e$ ) has certain weight that represents the frequency of joint authorships i.e. papers written jointly by a pair of authors.

The paper is organized as follows: In section 2, we discuss background and related work in the area. We discuss our data collection in section 3. In section 4, we discuss social network extraction and visualization. In section 5, we discuss social network analysis metrics we concentrate on in this work. We present and discuss our experimental results in section 6. Finally, we conclude and give some future directions in section 7.

## 2. RELATED WORK

Automatic techniques for social network extraction are not that old with the first one proposed in the year 1997 [10]. It was the first attempt of its kind to develop an automated interactive tool for extraction of social networks formed by

---

<sup>1</sup> Orgnet.com: <http://www.orgnet.com/sna.html>

people of a specific domain. Since then a number of such efforts have been made to automatically extract underlying social networks from a multitude of data sources. A detailed discussion on the social network extraction methods has been provided in [11]. These methods have been classified in [11] on the basis of the type of the information source used for extraction of relationships and in turn the social networks.

Co-occurrence of names on the web returned by a search engine in response to a query has commonly been used by social network extraction methods [10, 12, 13] to quantify the strength of relation between two names. Referral Web [10] extracts egocentric social networks by using co-occurrence of any two names in publically available online documents like homepages, publications, citations, etc. Domain specific social network extraction has been performed by some studies [13, 14], where [14] extracts social network formed by conference participants and [13] extracts social networks of online Semantic Web community.

In addition to co-occurrence based academic social networks some studies like [15] uses this measure to extract social networks from news articles Majority of the co-occurrence based methods proposed in the literature used co-occurrence based metrics to compute the weight of the extracted relations among entities but few of them have examined how to weigh each relation among entities beyond the co-occurrence based metrics [12]. Oka and Matuso [16] propose a method for weighting the relation among entities based on the weight of relations through the keyword, overcoming the shortcomings of the co-occurrence based metrics. The method receives a pair of entities and various relations that exist between entities as input. The output is the weight value for the pair of entities according to the generality of the keyword as a measure of its web hit counts.

There are a number of studies which concentrate on extraction of various types of social networks from different types of online data sources. Because of space constraints we cannot discuss all of them here but readers interested in the same can refer to [11].

The focus of the techniques discussed earlier in this section is extraction of social networks however they do not provide any analysis of these networks. In essence these networks are a result of some collaboration between the entities involved and their study from a network perspective may unravel some interesting facts about their structure, flow of information, important actors in the network, etc.

In addition to extraction of social networks some studies have tried to analyze and predict the evolution of collaboration networks in the scientific domain [5]. Some of them [17, 18, 19, 20] have tried to investigate their behavior in terms of their being small-world, scale-free, following power-law, etc. in addition to indentifying important actors in the network through parameters like betweenness centrality, vertex centrality etc. Analysis of these parameters provides a good insight of the health of the research community and the institution [5]. Better the academic and research activities in a group, community or an institution, the better their health.

Co-authorship being the most tangible and documented form of academic collaboration has the potential of being a true representative of academic collaborations. Thus co-authorship network can be considered as a true representative of academic social networks of people involved. Some studies [17, 18, 19, 20, 21, 22, 23, 24] have analyzed co-authorship networks across several domains like Biology, Computer

Science, Mathematics, Physics, Social Science, Database, Digital Library, etc.

Study of the cooperation through co-authorship relationship using social network analysis measures has been made in specific domains like journals and conferences. Networks formed in Chinese humanities and social science [25] in DBLP listed conferences viz. KDD, VLDB, ICML and WWW [26], in specific venues like IC3 [27] in Scientometrics Journal [28], etc.

### 3. DATA

The data for the purpose of this study are mined primarily from websites of four IITs. These IITs are IIT Kanpur, IIT Delhi, IIT Kharagpur and IIT Madras. We extracted faculty list and publications of each of these faculty members who are currently on the roll of Computer Science Engineering departments of these four IITs as full time faculty. The period of investigation has been restricted for five years from 2011 to 2015. Publications data of these faculty members were extracted either from their homepages or from some indexing service. Wherever, this data was not available on their homepages directly, it was mined from DBLP. This data was not in a condition that it could have been used directly for analysis purposes. It was first cleaned and data from all the sources brought in a common format so as to make it useful for any further processing.

In all we extracted 1082 publications of 111 faculty members published from 2011 onwards. The statistics of the dataset for the corresponding period are listed in Table-1.

**Table 1. Statistics of the dataset.**

| IIT          | Faculty    | Number of Publications in |            |            |            |           | Total       |
|--------------|------------|---------------------------|------------|------------|------------|-----------|-------------|
|              |            | 2011                      | 2012       | 2013       | 2014       | 2015      |             |
| Kanpur       | 25         | 17                        | 42         | 27         | 14         | 3         | 103         |
| Delhi        | 27         | 62                        | 71         | 72         | 69         | 14        | 288         |
| Kharagpur    | 32         | 40                        | 69         | 70         | 20         | 1         | 200         |
| Madras       | 27         | 113                       | 116        | 126        | 108        | 28        | 491         |
| <b>Total</b> | <b>111</b> | <b>232</b>                | <b>298</b> | <b>295</b> | <b>211</b> | <b>46</b> | <b>1082</b> |

### 4. SOCIAL NETWORK VISUALIZATION AND ANALYSIS

In order to convert joint publications into a collaboration network, the publications data has to go through a number of steps. Algorithm-1 provides an overview of the major steps involved in the process of transforming raw publication data into social network graphs. The same algorithm can be repeated for any number of institutes.

From an abstract point of view the steps involved in the social network extraction process can be viewed as shown in Figure-1. After data cleaning name disambiguation is performed in order to remove duplicates, if any. From these cleaned co-authors list co-author relations are extracted. We implemented this algorithm in Java.

Once these publications were brought in a common format, list of all the authors for each of the publications were extracted. Before extraction of co-authorship relationships from these author lists duplicate names has to be removed

otherwise we may not be able to either visualize or analyze these co-authorship networks correctly. In addition, these duplicates may also hamper the analysis of the research productivity of individual researchers. Removal of duplicates or name disambiguation plays an important role in efficient analysis of publications data. In case of academic social network analysis name disambiguation plays a very crucial role. A detailed discussion on name disambiguation techniques has been provided in [29]. We use a modification of vector space model based name disambiguation technique provided in [30].

**Algorithm-1: Social Network Extraction and Visualization.**

- Step-1 Extract faculty list from homepage of the department.
- Step-2 *for each* faculty extract publications from his homepage or DBLP
- Step-3 Clean and normalize these publications
- Step-4 *for each* publication in the publications list extract the list of authors
- Step-5 Perform name disambiguation and extract co-authorship relationships from these publications
- Step-6 Extract social networks from these co-authorship relationship
- Step-7 Visualize and analyze these social networks using a network graph visualization engine.

After performing name disambiguation, co-authorship relationships from each of the author lists were extracted. These relations were exported to NodeXL. NodeXL converts these co-authorship relationships into network graphs which render themselves to visual analysis. Social network metrics discussed in [5] and some others were obtained from graph metric calculation facility provided in NodeXL.

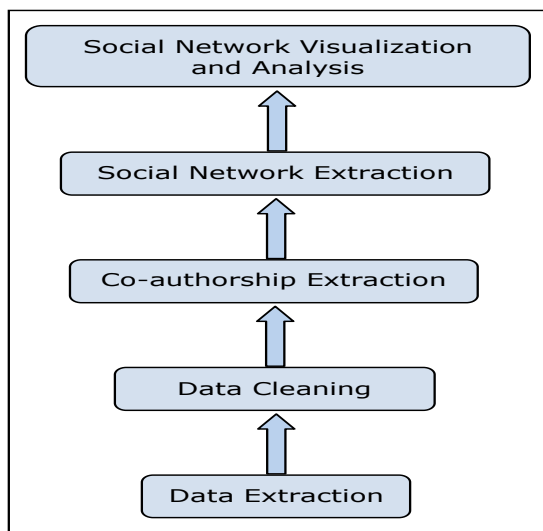


Figure 1: Workflow of the proposed social network extraction and visualization system.

## 5. SOCIAL NETWORK ANALYSIS METRICS

According to [31] the levels of social network analysis are actor, dyadic, triadic, subset, or network. Metrics like

centrality, prestige and roles such as isolates, liaisons, bridges, etc. are used to analyse the social network at actor level, whereas distance and reachability, structural and other notions of equivalence, and tendencies toward reciprocity are important at dyadic level. At triadic level one is interested in balance and transitivity. At subset level one is interested in finding cliques, cohesive subgroups, components whereas metrics like connectedness, diameter, centralization, density, prestige, etc. are used for analysis at network level [32].

Social network metrics, such as, centrality measures [33] (degree centrality, betweenness centrality, closeness centrality and network centrality), average degree, clustering coefficient, Salton index, density and characteristic path length are often of great interest to the analysis of academic or research collaboration from a network perspective. The study of social networks and their associated metrics is important as these networks form the underlying structure, which allows for rapid information distribution [34].

Several network analysis measures as proposed in [35] can be used to identify influential nodes and discover community structures of the extracted social networks. We are interested in capturing the internal connectivity as well as attributes of key nodes in the network. In order to identify the leaders in the network, the quantity of interest in many social network studies is the “betweenness centrality” of an actor ‘i’. Centrality is a measure of the relative importance of nodes and edges in a graph [33]. Several centrality measures like “betweenness centrality”, “closeness centrality” and “degree centrality” have been proposed in [35] to identify the most important actors (leaders) in a social network.

The objective of this study is to identify important actors like hubs/leaders, author having most number of connections, strength of collaboration ties, connectivity of authors, etc. using social network analysis metrics.

## 6. RESULTS AND DISCUSSIONS

The institutional networks presented in this section provide an insight into the amount of research activity being carried out by them. Simple metrics, like number of papers published by faculty in an institution, give some idea about the amount of research activity being carried out. However, advanced metrics like clustering coefficient provide an idea about the connectivity between various actors and cohesiveness of the network.

Table-2 lists the values of various network metrics for the co-authorship based social networks of IIT-Delhi, IIT-Kanpur, IIT-Kharagpur and IIT-Madras presented in Figures 2, 3, 4 and 5, respectively.

These values have been obtained using Graph Metrics Calculator available in the NodeXL Microsoft Excel Template [36]. NodeXL is an open source interactive network visualisation and analysis tool that uses MS Excel as the platform for visualization and analysis of networks using relationship data.

The values of various metrics listed in Table 2 provide important information, inter alia, on the structure of the network, connectivity in the network and patterns of communication. For example, from the analysis of the values of Average Clustering Coefficient listed in this table, it can be observed that it is highest for IIT-Delhi followed by IIT-Kharagpur. This also gets verified from the value of Connected Components for these two IITs. All these four collaboration networks exhibit Small World behavior because



the degree-of-separation (Average Geodesic Distance in Table-2) in all the cases is less than six.

Although visual picture may present IIT-Madras as most dense but IIT-Kharagpur has highest density followed by IIT-Kanpur. IIT Delhi has the lowest density of all the four graphs.

Table-2: Values of various metrics for co-authorship networks of the four IITs.

| Metrics                                | IIT     |        |           |         |
|--|---------|--------|-----------|---------|
|  | Delhi   | Kanpur | Kharagpur | Madras  |
| Vertices                               | 399     | 144    | 243       | 567     |
| Total Edges                            | 1949    | 312    | 1232      | 1982    |
| Maximum Geodesic Distance              | 7       | 4      | 10        | 11      |
| Average Geodesic Distance              | 3.9513  | 1.7417 | 4.0214    | 5.0033  |
| Graph Density                          | 0.0015  | 0.0235 | 0.0239    | 0.0082  |
| Maximum Degree                         | 61      | 23     | 34        | 90      |
| Average Degree                         | 6.085   | 3.361  | 5.811     | 4.667   |
| Max. Betweenness Centrality            | 21825.4 | 300.2  | 5400.0    | 80320.6 |
| Avg. Betweenness Centrality            | 503.80  | 5.125  | 324.63    | 897.282 |
| Maximum Closeness Centrality           | 0.2     | 1.0    | 0.5       | 1.0     |
| Average Closeness Centrality           | 0.004   | 0.159  | 0.01      | 0.01    |
| Max. Eigenvector Centrality            | 0.05    | 0.123  | 0.035     | 0.076   |
| Avg. Eigenvector Centrality            | 0.003   | 0.007  | 0.004     | 0.002   |
| Maximum PageRank                       | 7.364   | 4.759  | 6.448     | 17.915  |
| Connected Components                   | 3       | 20     | 3         | 7       |
| Max. Vertices in a Connected Component | 368     | 28     | 228       | 503     |
| Max. Edges in a Connected Component    | 1781    | 70     | 1162      | 1754    |
| Average Clustering Coefficient         | 0.859   | 0.766  | 0.807     | 0.775   |

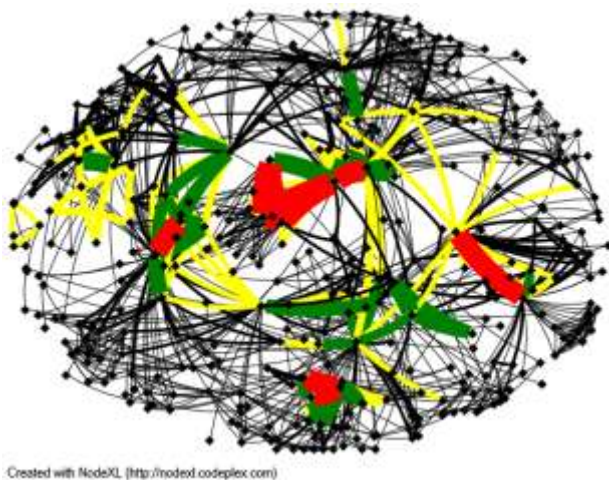


Figure 2: Co-authorship network of IIT-Delhi

In case of IIT-Delhi Aaditeshwar Seth has highest Degree followed by Vinay J. Ribero, Kolin Paul, Smruti R. Sarangi

and Amit Kumar. Amit Kumar has highest Betweenness Centrality followed by Vinay J. Ribero, Smruti R. Sarangi, Aaditeshwar Seth and Amitabha Bagachi. In addition to being the node with highest Degree Aaditeshwar Seth has highest Eigenvector Centrality as well as highest PageRank. Although Amit Kumar acts as bridge in the network Aaditeshwar Seth is highly connected with strong connections.

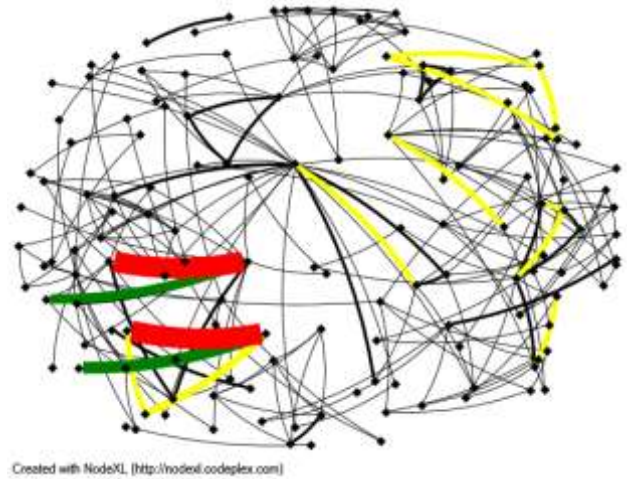


Figure 3: Co-authorship network of IIT-Kanpur

Arnab Bhattacharya has highest Degree in collaboration graph of IIT-Kanpur. He is followed by Phalguni Gupta, Surender Baswana, Harish Karnick and Amey Karkare in terms of Degree. Arnab Bhattacharya has highest Betweenness Centrality also followed by Akshay Mittal, Surender Baswana, Subhajit Roy and Harish Karnick. Arnab Bhattacharya also enjoys the status of having connections with other influential people in the network as he has highest Eigenvector Centrality as well as highest PageRank. Values of these metrics indicate that Arnab Bhattacharya acts as bridge in the network, has most number of connections and has strong connections in the network.

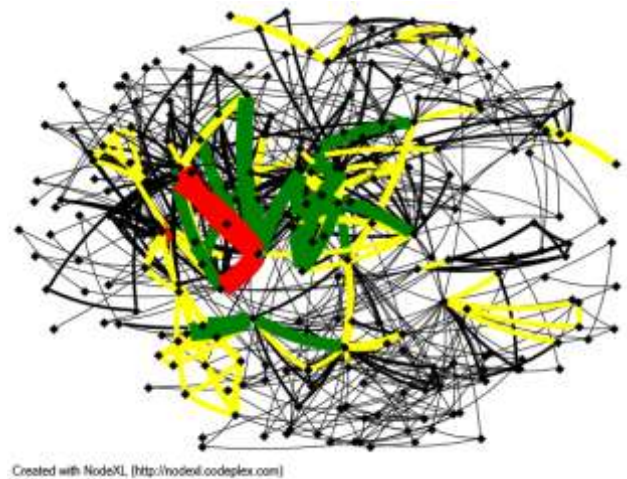


Figure 4: Co-authorship network of IIT-Kharagpur

In collaboration network of IIT-Kharagpur five persons having highest Degree are Pallab Dasgupta, Niloy Ganguly, Rajib Mall, J. Mukhopadhyay and J. Mukherjee respectively. Ansuman Banerjee has highest Betweenness Centrality followed by Rajib Mall, Sukanta Bhattacharya, Santosh Ghosh and Pallab Dasgupta. A. K. Majumdar has highest

Eigenvector Centrality but Rajib Mall has highest PageRank. This implies that Pallab Dasgupta has highest number of connections but A. K. Majumdar has strong connections i.e. connections with influential people in the network. Ansuman Banerjee acts as bridge in the network and Rajib Mall commands highest prestige.

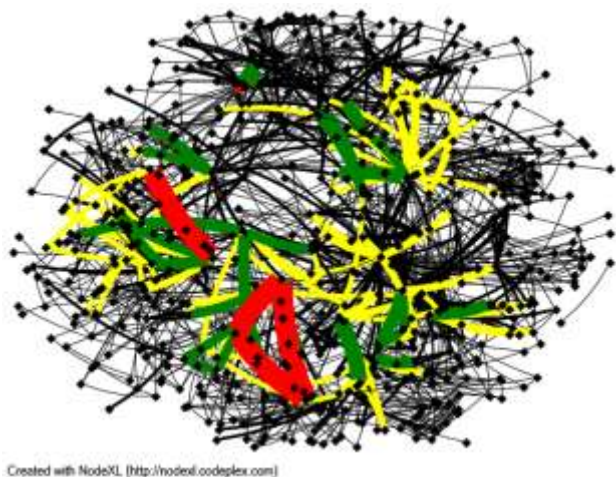


Figure 5: Co-authorship network of IIT-Madras

In this case Balam Ravindran has highest Degree followed by Krishna M. Sivalingam, C. Pandu Rajgan, N. S. Narayanaswamy and C. Siva Rama Murthy. Similarly Balam Ravindran has highest Betweenness Centrality. He is followed by N. S. Narayanaswamy, Sutanu Chakarborty, C. Pandu Rajgan and C. Siva Rama Murthy. C. Chandra Sekhar has strongest connections as he has maximum Eigenvector Centrality whereas Balam Ravindran enjoys highest prestige in the network as he has highest PageRank. This implies that Balam Ravindran has most number of connections in addition to being the bridge of the network.

In all of the four collaborations graphs differently coloured edges have been used. Red indicates strongest ties followed by Green, Yellow and Black. Black indicates weak ties. The width of the edges also complements the colour edges used in exhibiting the strength of collaboration ties.

## 7. CONCLUSIONS & FUTURE DIRECTIONS

Social network analysis has been used quite often in the past to study patterns of interaction among various entities including academic ones. In this work we extracted and analyzed co-authorship based collaboration networks of four leading IITs. It can be concluded that IIT Madras has been leading all the four IITs in terms of number of papers produced, per-capita productivity, etc. IIT Khargapur has a different network structure as compared to the rest of three as four different actors enjoy highest values for four different metrics on the basis of which we compare and contrast these collaboration networks.

As a part of future work we would like to extract social network on yearly basis and study the evolution of collaboration networks. We can also extract and analyze local collaborations between people from within a

particular department. Another direction could be to study the collaborations between these four institutes.

## 8. REFERENCES

- [1] Fernandes, J. M. (2014) Authorship trends in software engineering. *Scientometrics* 101: 257-271. doi:10.1007/s11192-014-1331-6
- [2] Greene, M. (2007) The demise of the lone author. *Nature*, 450(7173): 1165. doi:0.1038/4501165a.
- [3] Kennedy, D. (2003) Multiple authors, multiple problems. *Science*, 301(5634): 733. doi:10.1126/science.301.5634.733
- [4] Fu, T. Z. J., Song, Q. and Chiu, D. M. (2014) The academic social network. *Scientometrics* 101: 203-239. doi: 10.1007/s11192-014-1356-x
- [5] Arif, T., Ali, R. and Asger, M. (2012) Scientific co-authorship social networks: a case study of computer science scenario in India. *International Journal of Computer Applications*, 52(12): 38-45. doi:10.5120/8257-1790
- [6] Wagner, C. S. and Leydesdorff, L. (2005a) Mapping the network of global science: comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, 1(2): 185–208.
- [7] Wagner, C. S. and Leydesdorff, L. (2005b) Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10): 1608–1618.
- [8] Wagner, C. S. and Leydesdorff, L. (2008) International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4): 317–325.
- [9] Luo, Y.L. and Hsu, C.H. (2009) An empirical study of research collaboration using social network analysis. In *Proceedings of 2009 IEEE International Conference on Computational Science & Engineering*, Vancouver, Canada, 2009: 921-926.
- [10] Kautz, H., Selman, B., and Shah, M. (1997) The hidden web. *American Association for Artificial Intelligence magazine*, 18(2): 27–35.
- [11] Arif, T., Ali, R. and Asger, M. (2014a) Social network extraction: a review of automatic techniques. *International Journal of Computer Applications*, 95(1): 16-23. doi:10.5120/16558-3964
- [12] Matsuo, Y., Mori, J., and Hamasaki, M. (2006) POLYPHONET: An advanced social network extraction system from the web. In *Proceedings of the 15<sup>th</sup> International Conference on World Wide Web- WWW'06*, Edinburgh, Scotland, May 2006: 397-406.
- [13] Mika, P. (2005) Flink: Semantic web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3(2): 211-223.

- [14] Tomobe, H., Matsuo, Y. and Hasida, K. (2003) Social network extraction of conference participants. In Proceedings of 12<sup>th</sup> International Conference on World Wide Web- WWW'03, Budapest, Hungary, May 2003.
- [15] Pouliquen, B. and Atkinson, M. (2008) Extracting and learning social networks out of multilingual news. In Proceedings of the Social Networks and Application Tools Workshop-SocNet-08, Slovakia, 2008: 13-16.
- [16] Oka , M. and Matsuo, Y. (2009) Weighting relations in social networks using the web. In Proceedings of 23<sup>rd</sup> Annual Conference of the Japanese Society for Artificial Intelligence, Takamatsu, Japan, 2009: 1-2.
- [17] Newman, MEJ. (2001a) Scientific collaboration Networks-II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(016132).
- [18] Newman, MEJ. (2001b) The structure of scientific collaboration networks. In Proceedings of the National Academy of Sciences, January, 2001, 98(2): 404-409.
- [19] Newman, MEJ. (2004) Co-authorship networks and patterns of scientific collaboration. In Proceedings of the National Academy of Sciences, 101(90001):5200-5205.
- [20] Newman, MEJ. (2001c) Scientific collaboration networks-I. Network construction and fundamental results. *Physical Review E*, 64(1):16131.
- [21] Moody, J. (2004) The structure of a social science collaboration network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review*, 69(2): 213-238.
- [22] Liu, X., Bollen, J. Nelson, M.L. and Van de Sompel. H. (2005) Coauthorship networks in the digital library research community. *Information Pro-cessing & Management*, 41(6): 1462-1480.
- [23] Sharma, M. and Urs, S.R. (2008) Network Dynamics of Scholarship: A Social Network Analysis of Digital Library Community. In Proceeding of the 2<sup>nd</sup> PhD Workshop on Information and Knowledge Management, New York, NY, USA, 2008: 101-104.
- [24] Elmacioglu, E. and Lee, D. (2005) On six degrees of separation in DBLP-db and more. *ACM SIGMOD Record*, 34(2): 33-40.
- [25] Ma, F., Li, Y. and Chen, B. (2014) Study of the collaboration in the field of the Chinese humanities and social sciences. *Scientometrics*, 100(2): 439-458. doi:10.1007/s11192-014-1301-z
- [26] Coscia, M., Giannotti, F. and Pensa, R. (2009) Social network analysis as knowledge discovery process: A case study on digital bibliography. In Proceedings of 2009 Advances in Social Network Analysis and Mining, Athens, Greece, 2009: 279-283.
- [27] Arif, T., Asger, M., Malik, M.B. and Ali, R. (2015) Extracting Academic Social Networks Among Conference Participants. In Proceedings of 8<sup>th</sup> International Conference on Contemporary Computing- IC3-2015, Noida, India, IEEE Press, 2015: 42-47.
- [28] Hou, H., Kretschmer, H., & Liu, Z. (2008) The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75(2): 189–202. doi:10.1007/s11192-007-1771-3
- [29] Ferreira , A. A., Gonçalves M. A., and Laender, A.H.F. (2012) A brief survey of automatic methods for author name disambiguation. *ACM SIGMOD Record*, 41(2): 15-26. doi:10.1145/2350036.2350040
- [30] Arif, T., Ali, R. and Asger, M. (2014b) Author name disambiguation using vector space model and hybrid similarity measures. In Proceedings of 7<sup>th</sup> International Conference on Contemporary Computing- IC3-2014, Noida, India, IEEE Press, 2014: 135-140. doi:10.1109/IC3.2014.6897162
- [31] Wasserman, S. and Faust, K. (1994) *Social network analysis: Methods and applications*. Cambridge University Press, New York, U.S.A.
- [32] Gretzel, U. (2001) *Social network analysis: Introduction and resources*. <http://irs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/#analysis>: Last Accessed: January, 2015.
- [33] Newman, M.E.J. (2010) *Networks: An Introduction*. Oxford University Press, United Kingdom.
- [34] Newman, M.E.J. (2006) Community centrality. *Phys. Reviews*, 74.
- [35] Chelmiss, C. and Prasanna, V.K. (2011) Social networking analysis: A state of the art and the effect of semantics. In Proceedings of 3<sup>rd</sup> IEEE Conference on Social Computing, Boston, MA, USA, 2011: 531-536.
- [36] Hansen, D. L., Shneiderman, B. and Smith, M. A. (2011) *Analyzing Social Media Networks with NodeXL*. Morgan Kaufman, 2011



# Analyzing Customer Behaviour through Data Mining

Sandeep Kumar  
Department of Computer Science  
Krishna Engineering College  
Ghaziabad, UP, India

Rakesh Kumar Arora  
Department of Computer Science  
Krishna Engineering College  
Ghaziabad, UP, India

---

**Abstract:** Indian organized retail industry is poised for growth. In order to attract customers price discounts in retail sector are the norm rather than an exception. The most common type of monetary promotions includes discounts, coupons, and rebates. Promotions helps organizations to grow market share, increase sales volume, sell faster, cultivate loyal customers, and drown out competitor advertising. With the widespread use of sales promotions, it has become important for the managers to understand such practices and understand challenges.

This paper will assist the manager in identifying the set of customers that are attracted towards the departmental store when the discount coupons were issued. The manager will also be able to formulate new policies to attract new set of customers.

**Keywords:** Data Mining, Retail Sector, Decision Tree, WEKA

---

## 1. INTRODUCTION

A large number of retail chains have opened over the last decade with the objective of providing quality products at low prices to the customers. Fierce competition and narrow profit margins have pushed retailers in implementing data warehouse earlier than other industries. This has given the retailers a better opportunity to take advantage of data mining.

Large retail chains and grocery stores store vast amounts of data collected over the years that are rich in information. Data Mining helps in reducing information overload along with the improved decision-making by searching for relationships and patterns from the huge dataset collected by retailers. It enables a retail industry to focus on the most important information in the database and allows retailers to make more knowledgeable decisions by predicting future trends and behaviors. The Data Mining uses the business data as raw material using a predefined algorithm to search through the vast quantities of raw data, and group the data according to the desired criteria that can be useful for the future target marketing.[1]

This paper uses Data Mining Technique to improve the sales in the departmental store by distribution of coupons among customers visiting the departmental store such that both customers and departmental stores can gain because of increased sales volume. Data mining, the extraction of hidden predictive information from large databases is a powerful technology with great potential to help managers in the departmental stores to have larger market share and cultivate loyal customers. It discovers information within the data that queries and reports can't effectively reveal. After gathering data regarding customer profiles submitted

by customers at the time of issuing of coupons and feedback form filled by the customers at the time of redemption of coupons, data mining technique need to be applied to identify set of customers that will help in increase sales volume and market share.

With the help of data mining techniques, such as clustering, decision tree or association analysis it is possible to discover the key characteristics from the details of customers and possibly use those characteristics for future prediction. This paper presents decision tree algorithm as a simple and efficient tool to analyze the customer details and distribution of coupons for higher sales and larger market share.[2]

## 2. METHODOLOGY

Decision trees are a simple, but powerful form of multiple variable analysis. A decision tree is a special form of tree structure. The tree consists of internal nodes where a logical decision has to be made, and connecting branches that are chosen according to the result of this decision. The nodes and branches that are followed constitute a sequential path through a decision tree that reaches a leaf node (final decision) in the end.[3]

In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most frequent case, each test considers a single attribute, such that the instance space is partitioned according to the attribute's value. In the case of numeric attributes, the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value. [4]

The decision tree algorithm is simple top down greedy algorithm. The major step of algorithm is to continue to divide leaves that are not homogeneous into leaves that are as homogeneous as possible until no further division is possible. The algorithmic steps for decision tree algorithm is as follows:[5]

1. Let the set of training data be S. If some of the attributes are continuous-valued, they should be discretized. Once that is done, put all of S in single tree node.
2. If all the instances in S are in same class, then stop.
3. Split the next node by selecting an attribute A from amongst the independent attributes that best divides or splits the objects in the node into subsets and create decision tree node.
4. Split the node according to the values of A
5. Stop if any of the following conditions are met, otherwise continue with step 3

**Figure 1: Steps for Decision Tree Algorithm**

Pruning is very important technique to be used in tree creation because of outliers. It also addresses overfitting. Datasets may contain little subsets of instances that are not well defined. To classify them correctly, pruning can be used. There are two types of pruning:

1. Post pruning (performed after creation of tree)
2. Online pruning (performed during creation of tree) [6].

The steps to extract classification rules from tree are mentioned below:

1. Represent the knowledge in the form of IF-THEN rules.
2. One rule is created for each path from the root to a leaf.
3. Each attribute-value pair along a path forms a conjunction.
4. The leaf node holds the class prediction

The analysis using decision tree is being done with the help of WEKA tool. WEKA, formally called Waikato Environment for Knowledge Learning supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. WEKA is an open source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java and is compatible with almost every computing platform. It is user friendly with a graphical interface that allows for quick set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type,

normal alpha-numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces. [7]

### 3. ANALYSIS

The study was carried out on the profiles of the customers who have visited the departmental store from July 2014 to Dec 2014. The attributes considered for analysis of customers along with their description are reflected in Table 1.

| Parameters      | Description  |
|-----------------|--|
| Sex             | Male (M) /Female (F)   |
| Age_Group       | Less than and equal to 20(Y) / More than 20 and less than equal to 30(L) / More than 30 and less than equal to 40(M) / More than 40 and less than equal to 50(N) / Above 50(O) |
| Profession      | Salaried (S) / Businessman (B)   |
| Qualification   | Under graduate (U) / Graduate (G) / Post graduate (P)  |
| Income          | Less than and equal to 25,000/- (Low) / More than 25,000 and less than 50,000/- (Medium) / More than 50,000/- (High)   |
| Coupon Utilized | Yes / No   |

**Table 1: Parameters used for analysis**



|    | A   | B         | C          | D             | E          | F               |
|----|-----|-----------|------------|---------------|------------|-----------------|
| 1  | Sex | Age_Group | Profession | Qualification | Income(Rs) | Coupon Utilized |
| 2  | M   | Y         | S          | U             | Low        | Y               |
| 3  | F   | L         | S          | G             | Low        | Y               |
| 4  | F   | M         | B          | P             | Low        | Y               |
| 5  | M   | L         | S          | U             | Low        | Y               |
| 6  | M   | O         | B          | G             | Low        | N               |
| 7  | F   | O         | S          | P             | Low        | N               |
| 8  | F   | M         | B          | U             | Low        | N               |
| 9  | M   | L         | S          | G             | Medium     | Y               |
| 10 | F   | M         | B          | P             | Medium     | N               |
| 11 | M   | Y         | S          | U             | Low        | Y               |
| 12 | M   | M         | B          | G             | Medium     | N               |
| 13 | M   | L         | S          | P             | Low        | Y               |
| 14 | F   | L         | B          | U             | Low        | N               |
| 15 | F   | L         | S          | G             | Low        | N               |
| 16 | F   | M         | B          | P             | Low        | Y               |
| 17 | M   | M         | S          | U             | Low        | Y               |
| 18 | M   | N         | B          | G             | Medium     | N               |
| 19 | M   | N         | S          | P             | Low        | Y               |
| 20 | M   | Y         | B          | U             | Low        | Y               |
| 21 | F   | O         | S          | G             | Low        | Y               |
| 22 | F   | O         | B          | P             | Low        | N               |
| 23 | F   | Y         | S          | U             | Low        | N               |

Figure 2: Sample view of Dataset

The data file normally used by WEKA is in ARFF (Attribute-Relation File Format) file format, which consist of special tags to indicate different things in the data file. Figure 2 shows the sample view of dataset and Figure 3 shows the ARFF format of desired dataset. To convert an Excel format into ARFF format an Excel to ARFF convertor is being used. The ARFF format dataset is represented in Figure 3

```

RETAIL - Notepad
File Edit Format View Help
@relation sheet3

@attribute Sex { M,F }
@attribute Age_Group { Y,L,M,O,N }
@attribute Profession { S,B }
@attribute Qualification { U,G,P }
@attribute Income(Rs) { Low,Medium,High }
@attribute Coupon { Y,N }

@data
M,Y,S,U,Low,N
F,L,S,G,Low,Y
F,M,B,P,Low,Y
M,L,S,U,Low,N
M,O,B,G,Low,N
F,O,S,P,Low,N
F,M,B,U,Low,N
M,L,S,G,Medium,N
F,M,B,P,Medium,N
M,Y,S,U,Low,Y
M,M,B,G,Medium,N
M,L,S,P,Low,Y
F,L,B,U,Low,N
F,L,S,G,Low,N
F,M,B,P,Low,Y
M,M,S,U,Low,Y
M,N,B,G,Medium,N
M,N,S,P,Low,Y
M,Y,B,U,Low,Y
F,O,S,G,Low,N
F,O,B,P,Low,N
F,Y,S,U,Low,N
F,M,B,G,Low,N
F,M,S,P,Low,N
M,M,B,U,Low,N
M,M,S,G,Low,Y
M,M,B,P,Low,N
    
```

Figure 3: ARFF Format of Sample Dataset

After collecting and cleaning the data, the classification of data is done using J48. J48 is an open source Java implementation of the C4.5 algorithm in the WEKA data mining tool. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The output generated is displayed in Figure 4.



Figure 4: Output

The accuracy is around 58%. The kappa statistic measures the agreement of prediction with the true class where value 1.0 signifies complete agreement. The confusion matrix or contingency table in this example has thirteen classes, and therefore a 13x13 confusion matrix is being displayed. The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified. The True Positive (TP) rate is the proportion of examples which were classified as class x, among all examples which truly have class x, i.e. how much part of the class was captured. It is equivalent to Recall. The False Positive (FP) rate is the proportion of examples which were classified as class x, but belong to a different class, among all examples which are not of class x. The Precision is the proportion of the examples which truly have class x among all those which were classified as class x. The F-Measure is simply  $2 * Precision * Recall / (Precision + Recall)$ , a combined measure for precision and recall.[8]

As per J48 Algorithm, parameters that reflect noise or outliers need to be removed, hence only those targeted node are shown by tree which have some value of precision and recall. The tree generated is represented in Figure 5.

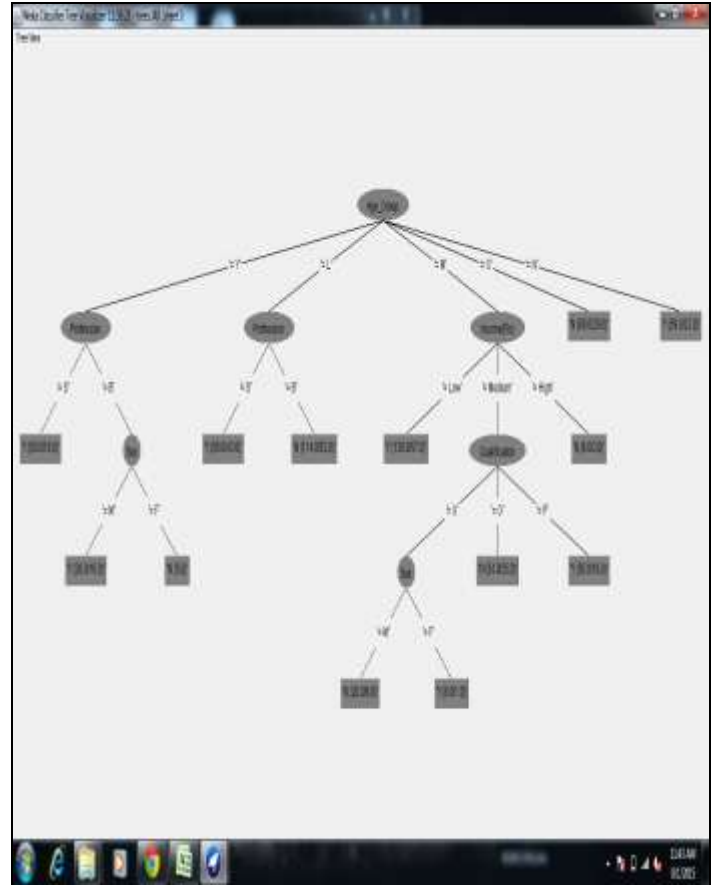


Figure 5: Decision Tree

## 5. RESULTS

The classification rules extracted from tree when the coupon is utilized are:

- 1) If Age\_group is less than 20 and Profession is Service than coupon is utilized.
- 2) If Age\_group is less than 20 ,Profession is Businessman and Sex is Male than coupon is utilized.
- 3) If Age\_group is between 20 and 30 and Profession is Service than coupon is utilized.
- 4) If Age\_group is between 30 and 40 , Income is than equal to Rs 25,000/- than coupon is utilized.
- 5) If Age\_group is between 30 and 40 , Income is more than Rs 25,000/- and less than Rs 50,000/- , Qualification is under graduate and Sex is Female than coupon is utilized.
- 6) If Age\_group is between 30 and 40 , Income is more than Rs 25,000/- and less than Rs 50,000/- and Qualification is post graduate than coupon is utilized.

- 7) If Age\_group is between 40 and 50 than coupon is utilized.

**Disadvantages of J48 algorithm:** The run-time complexity of the algorithm matches to the tree depth, which cannot be greater than the number of attributes. Tree depth is linked to tree size, and thereby to the number of examples. So, the size of C4.5 trees increases linearly with the number of examples. C4.5 rules slow for large and noisy datasets Space complexity is very large as we have to store the values repeatedly in arrays [9].

## 6. CONCLUSION

In this paper, a simple methodology based on decision tree algorithm is being used to analyze the customer details for distribution of coupons. This methodology will assist the Managers in the store to identify set of customers that are likely to purchase from store. The manager will also be able to devise the new schemes in order to attract new set of customers. This model will play important role in understanding demographics of customers by clearly differentiating between the customers that need to be retained and that need to be targeted. This will have significant effect in improving sales and hereby achieving targets of departmental store.

## 7. REFERENCES

- [1] Ahmed, S. R. (2004), 'Applications of Data Mining in Retail Business', Proceedings of the International Conference on Information Technology: Coding and Computing, Vol.2, pp. 455- 459 IEEE
- [2] Arora K. Rakesh, Badal Dharmendra, "Admission Management using Data Mining using WEKA", IJARCSSE Vol. 3, Issue 10, October 2013
- [3] [Online] [http://www.estard.com/decisiontree/decision\\_trees\\_definition.asp](http://www.estard.com/decisiontree/decision_trees_definition.asp)
- [4] [Online] <http://www.ise.bgu.ac.il/faculty/liorr/hbc/hap9.pdf>
- [5] Gupta K G. "Introduction to Data Mining with case studies", PHI
- [6] Moertini, Veronica S. "Towards the use of C4.5 algorithm for classifying banking dataset." Vol. 8 No. 2, October 2003 (2003): 12. Web. 24 Jan. 2013
- [7] [Online] <http://www.gtbit.org/downloads/dwdmsem6/dwdmsem6lman.pdf>
- [8] [Online] <http://weka.wikispaces.com/Primer>
- [9] Juneja, Deepti, et al. "A novel approach to construct decision tree using quick C4.5 algorithm." Oriental Journal of Computer Science & Technology Vol. 3(2), 305-310 (2010) (2010): 6. Web. 18 Feb. 2013.
- [10] Arora K. Rakesh, Badal Dharmendra, "Subject Distribution using Data Mining", IJRET Vol. 2, Issue 12, December 2013
- [11] Arora K. Rakesh, Badal Dharmendra, "Placement Prediction through Data Mining", IJARCSSE Vol. 4, Issue 7, July 2014.
- [12] [Online] <http://www.gallimoreinc.com/learn/how-coupons-help-sell-more-product.php>

# The Mathematics of Social Network Analysis: Metrics for Academic Social Networks

Tasleem Arif

Department of Information Technology  
Baba Ghulam Shah Badshah University  
Rajouri, J&K, India

---

**Abstract:** Social network analysis plays an important role in analyzing social relations and patterns of interaction among actors in a social network. Such networks can be casual, like those on social media sites, or formal, like academic social networks. Each of these networks is characterised by underlying data which defines various features of the network. Keeping in view the size and diversity of these networks it may not be possible to dissect entire network with conventional means. Social network visualization can be used to graphically represent these networks in a concise and easy to understand manner. Social network visualization tools rely heavily on quantitative features to numerically define various attributes of the network. These features also referred to as social network metrics used everyday mathematics as their foundations. In this paper we provide an overview of various social network analysis metrics that are commonly used to analyse social networks. Explanation of these metrics and their relevance for academic social networks is also outlined

**Keywords:** Social network analysis, mathematics, metrics, academic networks

---

## 1. INTRODUCTION

Social networks have been around us since time immemorial. They can be found around us in a variety of shapes and forms. They can be real or virtual but their basic properties still remain the same. Social network has been defined in different ways but their reliance on the fundamentals of mathematics has remained unchanged in almost all of the definitions. Social networks have been and are increasingly being represented through mathematical representations like graphs, matrices and relations. A standard definition of social network can be found in [1] as “a structured representation of the social actors (nodes) and their interconnections (ties)”. These networks can be represented as a graph  $G = (V, E)$ . The set  $V$  denotes entities (people, places, organizations, Webpages, etc.) joined in pairs by edges in  $E$  denoting acquaintances or relationships (friends, siblings, co-authors, hyperlinks, etc.) [2].

Social networks are ubiquitous characterized by underlying social groups that share common interests. These networks and the underlying groups have emerged on the Web at a rapid pace and have become one of the widely used online activity [3]. These networks are an aggregation of groups or virtual communities with each of these communities different from the other in composition, purpose and intent. Members of these virtual communities profit from being linked to other people sharing common interests despite their geographically dispersed affiliations. Social networks can be constructed for business entities like a company or firm, for educational entities like a school or University, or for any other set of entities [4].

Social networks have got a lot of attention from the research community long before the advent of the Web [5]. Between 1950 and 1980, when Vannevar Bush’s proposed hypertext medium ‘Memex’ was gaining acceptance, Social Sciences also contributed a lot in measuring and analyzing social networks [6]. There are numerous examples of social networks formed by social interactions like co-authoring, advising, supervising, and serving on committees between academics; directing, acting, and producing between movie

personnel; composing and singing between musicians; trading and diplomatic relations between countries; sharing interests, connections, and transmitting infections between people; hyper linking between Web pages; and citations between papers.

There have been a number of efforts to study these networks with the first formal attempt of its kind undertaken around eight decades ago. Manual methods can be used to analyze these networks but with new actor and relationship the complexity of the network increases many folds thus rendering manual methods ineffective. With the advent of Internet and developments in information and communication technologies the size, reach and diversity of these networks has become immense. One can’t even think of analyzing these networks with a simple computer leave aside manual techniques. One can’t even think of keeping aside these networks because they engage enormous number of users thus providing a large customer base to businesses on the one hand and the most common medium of interaction among geographically separated users on the other. To meet this demand a specialized science called Social Network Analysis (SNA) with its roots in social sciences particularly Sociology has emerged. With each passing day the dependence of these networks on mathematics and mathematical tools has been increasing.

The focus of Social Network Analysis (SNA) is relationships, their patterns, implications, etc. Using it, one can study these patterns in a structural manner [5]. SNA can be used to identify important social actors, central nodes, highly or sparsely connected communities and interactions among actors and communities in the underlying network [5]. SNA has been used to study social interaction in a wide range of domains, e.g. collaboration networks [7], directors of companies [8], inter-organizational relations [9], etc.

The study of social networks for behaviour analysis of actors involves two aspects: (a) the use of formal theory organized on the basis of mathematical conventions and (b) the empirical analysis of network data as quantified by various social network analysis metrics. So it can be understood that



social network metrics play an important role in SNA. This paper identifies various social network metrics and the mathematics behind them. These metrics have different meanings in different types of networks. In addition this paper also examines the use and relevance of these metrics in academic social networks.

## 2. SOCIAL NETWORK ANALYSIS: LEVELS AND METRICS

Like in other fields metrics help define certain attributes in quantitative terms. This section illustrates different levels of social network analysis along with the metrics that are used to draw inferences about the network.

There are five different levels of social network analysis, each of them characterised by the structure of the underlying network. It may be at actor level, dyadic level, triadic level, subset level, or network level. Metrics like centrality, prestige and roles such as isolates, liaisons, bridges, etc. are used to analyse the social network at actor level, whereas distance and reachability, structural and other notions of equivalence, and tendencies toward reciprocity are important at dyadic level. At triadic level one is interested in balance and transitivity. At subset level one is interested in finding cliques, cohesive subgroups, components whereas metrics like connectedness, diameter, centralization, density, prestige, etc. are used for analysis at network level<sup>1</sup>.

Some of the commonly used SNA metrics are:

**Centrality:** As said earlier ‘relationships’ is the focus of SNA and the ‘actors’ are central to all types of relationships. Thus attribute description or profiling of actors is an important aspect of any social network analysis. In this context Chelmiss and Prasanna [10] proposed several social network analysis measures (metrics) that can be used to identify influential nodes in a social networks. Centrality is a measure of the information about the relative importance of nodes and edges in a graph. Centrality measures like Degree Centrality, Closeness Centrality, Betweenness Centrality, Eigenvector Centrality, Katz Centrality and Alpha Centrality play an important role in graph theory and network analysis to measure the importance or prestige of actors or nodes in a network<sup>2</sup>. Several centrality measures like betweenness centrality, closeness centrality, and degree centrality have been proposed in [10] to identify the most important actors (leaders) in a social network.

- **Degree Centrality:** It is the simplest of all the centrality measures and its value for a given node in the network is the number of links incident on it and is used to identify nodes that have highest number of connections in the network. However it does not takes into account the centrality or prestige of the incident nodes. For a graph  $G = (V, E)$ , the degree of a node or vertex  $v, (v \in V)$  can be expressed using Equation (1).

$$C_D(v) = deg(v) \quad (1)$$

where  $deg(v)$  is the number of edges incident on the vertex  $v$ .

For entire graph  $G$  the *Degree Centrality* can be expressed using Equation. 2.

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v^*) - C_D(v_i)]}{H} \quad (2)$$

Where  $v^*$  is the node in  $G$  with highest degree centrality and  $H = \sum_{j=1}^{|V|} C_D(y^*) - C_D(y_j)$ , where  $y^*$  be the node with the highest degree centrality in a graph  $X$  of  $G$  with  $Y$  nodes. The value of  $H$  is maximum when a graph has a star like structure.

- **Eigenvector Centrality:** A more sophisticated version of degree centrality is eigenvector centrality. It not only depends on the number of incident links but also the quality of those links. This means that having connections with high prestige nodes contributes to the centrality value of the node in question. Google’s *PageRank* and *Katz Centrality* is a variation of eigenvector centrality and closely related to eigenvector centrality respectively.

Let  $A = (a_{v,u})$  be the adjacency matrix of a graph  $G$  with  $V$  vertices and  $E$  edges. Then  $A$  can be defined as:

$$A_{v,u} = \begin{cases} a_{v,u} = 1, & \text{if vertex 'v' is linked to vertex 'u'} \\ a_{v,u} = 0, & \text{otherwise} \end{cases}$$

The eigenvector centrality of a vertex  $v$  can be defined using Equation (3).

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in N(v)} x_u = \frac{1}{\lambda} \sum_{u \in G} a_{v,u} x_u \quad (3)$$

where  $N(v)$  represents the set of neighbours of the vertex  $v$  and  $\lambda$  is a constant.

- **Closeness Centrality:** The degree of nearness (direct or indirect) between any node and rest of the nodes in the network is represented by “closeness centrality”. It is the inverse of sum of the shortest distance (also called geodesic distance) between a node and rest of all in the network. For a graph  $G$  with ‘ $n$ ’ nodes the closeness centrality of a node ‘ $v$ ’ can be expressed using Equation (4).

$$C_C(v) = \frac{n - 1}{\sum_{k=i}^n d(u_i, v)} \quad (4)$$

where  $d(u_i, v)$  denotes the geodesic distance between  $u_i$  and  $v$ .

- **Betweenness Centrality:** In order to identify the leaders in the network, the quantity of interest in many social network studies is the “betweenness centrality” of an actor ‘ $i$ ’. Betweenness centrality measures the fraction of all shortest paths that pass through a given node or in simple terms it quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. Nodes with high betweenness centrality play a crucial role in the information flow and cohesiveness of the network and are considered central and indispensable to the network due to their role in the flow of information in the network. Nodes with the high betweenness act as

<sup>1</sup><http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/#analysis>

<sup>2</sup> <http://en.wikipedia.org/wiki/Centrality>

gate keeper. The betweenness centrality of vertex  $v$  can be expressed using Equation (5).

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (5)$$

where  $\sigma_{st}$  is the total number of shortest paths from node  $s$  to  $t$  and  $\sigma_{st}(v)$  is the number of paths that pass through  $v$ .

- **Clustering coefficient:** It signifies how well a node's neighbourhood is connected<sup>3</sup>. Clustering coefficient is a measure of the ability of a node's neighbour to form a complete graph, also called a clique. The value of clustering coefficient is directly proportional to the degree of connectedness of the neighbours of that node: more the connections among the neighbours, the higher the clustering coefficient. The clustering coefficient of a network as given in [11] is the average of the clustering co-efficient of all the nodes in the network. It is therefore considered to be a good measure if a network demonstrates "small world" behaviour [11]. Stanley Milgram's [12] theory of the "6 Degree of Separation" utilises the average path length metric. A graph is considered small world if its average clustering coefficient is significantly higher than a random graph constructed from the same set of vertices.

The average clustering coefficient can be expressed using Equation (6) as follows:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i \quad (6)$$

where  $C_i = \frac{\lambda_G(v)}{\tau_G(v)}$ ,  $\lambda_G(v)$  is the number of subgraphs of  $G$  having 3 edges and 3 vertices including the vertex  $v$ .  $\tau_G(v)$  is the number of subgraphs of  $G$  having 2 edges and 3 vertices including  $v$  such that  $v$  is incident on both the edges.

- **Average Degree:** The number of vertices adjacent to a vertex  $v$  is called as the degree of  $v$  or  $deg(v)$ . Based on this measure one can get maximum degree, minimum degree or average degree. The average degree of a graph is a network level measure and it is calculated from the value of degree or all the nodes in the network. For a graph  $G$  with  $V$  vertices and  $E$  edges the average degree of  $G$  can be expressed using Equation (7).

$$D_A(G) = \frac{2 \times |E|}{|V|} \quad (7)$$

- **Density:** The Density of a graph quantifies the number of connections between various actors in the network. The graph is considered dense if the number of edges in the graph approaches the maximal number of edges which one can have in that graph and sparse otherwise. For an undirected

graph  $G$  with  $V$  vertices and  $E$  edges, the density of  $G$  can be expressed using Equation (8) as follows:

$$D_G = \frac{2|E|}{|V|(|V| - 1)} \quad (8)$$

### 3. SNA METRICS FOR ACADEMIC SOCIAL NETWORKS

Universities, research laboratories and other institutions of higher learning are known for providing solutions to various problems confronting the society [13]. Research has been providing answers to many such problems. Modern day research is faced with both extraordinary opportunities and challenges. A fast paced modern society turns to academics for immediate answers to an array of practical problems created by its own increasing needs and desires. Society is willing to invest in research as the basis of a knowledge economy as long as research proves to be responsive to its needs, is productive and effective. Knowledge sharing and interactions are at the heart of research practice and collaboration. Collaboration is defined as "working jointly with others or together especially in an intellectual endeavor"<sup>4</sup>. Research interactions and collaborations include working on a research project jointly and publishing the results of the research undertaken. These collaborations help promote and proliferate research [13, 14], therefore, they should be encouraged, supported and monitored. Studies [13, 15, 16, 17] indicate that there is a direct relationship between scientific collaboration and creation of new knowledge. Co-authorship is one of these collaborations. In order to understand and analyse the social networks formed by any form of academic collaborations, they need to be viewed from a network perspective. SNA metrics can be used to seek answer, inter alia, to the following questions:

- Who are the hubs/leaders?
- Who has more connections?
- How strong are the collaboration ties?
- How collaborative the authors are?
- How connected the network is?

Value of various SNA metrics discussed above can be used to answer these and many other questions that help us understand the structure of network, flow of information in the network, strategic positions occupied by the authors in the network, important individuals, prestige of important authors in the network, etc. In the following we discuss the applicability of the above listed SNA metrics in academic social networks.

- **Degree Centrality:** In case of academic social networks degree centrality means the centrality of an actor in terms of frequency of the considered activity. The more the activity the better the degree centrality. For example in co-authorship networks it is a measure of how often an author collaborates with other authors in the network. However it does not takes into account the quality of collaborators. Having connections with such nodes (authors) may not necessarily rate you

<sup>3</sup> [http://en.wikipedia.org/wiki/Clustering\\_coefficient](http://en.wikipedia.org/wiki/Clustering_coefficient)

<sup>4</sup> Merriam-Webster's Collegiate Dictionary (1999). Tenth Edition. Springfield, MA: Merriam-Webster, Incorporated.

higher in terms of your prestige in the academic social network.

- **Eigenvector Centrality:** Since the value of eigenvector centrality of a node depends upon the quality of connections nodes with higher eigenvector centrality lie at the centre of flow of ideas and information in the network. In co-authorship networks it is a representation of an author's ability to receive new research ideas that spread across the network [18].
- **Betweenness Centrality:** Nodes with high betweenness centrality occupy strategic positions in the network. Removal of such nodes result in breakdown of the information flow and the nature of connectivity in the network may change altogether. Analytical results obtained [19] testify that in academic social networks actors (scientists in this case) having high value of betweenness centrality in a network play a positive role in advancing scientific cooperation.
- **Closeness Centrality:** It is measure of the proximity of an academic with others in the network. Here the diversity is link is important than the quality of links. If a node is connected with majority of other nodes in the network, either directly or indirectly, the closeness centrality of that node will be more than of those have connections with other high profile nodes.
- **Clustering Co-efficient:** Measure of connectivity in the network. In academic social networks clustering coefficient means is a way of predicting future collaborations between any two academics that are indirectly collaborating with each other i.e. collaborating through a mutual collaborator [20].
- **Average Degree:** Each of the nodes (academics) may have different potential of connectivity with other nodes in the network. It is a network metric and in academic social networks it is considered as a measure of how collaborative the academics are.
- **Density:** The density refers to the potential of connectivity in the network. In academic social networks it represents the degree of collaboration that takes place in the network [18].

#### 4. CONCLUSIONS

Mathematics has been called as mother of all the sciences and SNA is no exception. Fundamentals of mathematics play an important role in the formulation of SNA. A social network can have any shape and form but the basic considerations remain almost same. In this paper we explained various social network analysis metrics and their dependence on mathematical concepts. After elaborating these metrics we discussed their use and relevance in analysis of academic social networks.

#### 5. REFERENCES

- [1] Jin, Y., Matsuo, Y. and Ishizuka, M. (2006) Extracting a social network among entities by web mining. In Proceedings of ISWC'06 Workshop on Web Content Mining with Human Language Technologies, Athens, GA, USA.
- [2] Lee, J. (2007) A study of collaborative product commerce by co-citation analysis and social network analysis. In Proceedings of 2007 IEEE International

Conference on Industrial Engineering and Engineering Management, Singapore, pp. 209-213.

- [3] Mislove, A., Marcon, M., Gummadi, K. P. Drushel, P., and Bhattacharjee, B. (2007) Measurement and analysis of online social networks. In Proceedings of the 5th ACM/USENIX Internet Measurement Conference, San Diego, CA, USA, pp. 29-42.
- [4] Arif, T., Ali, R. and Asger, M. (2012) Scientific co-authorship social networks: A case study of computer science scenario in India. International Journal of Computer Applications, 52(12), pp. 38-45.
- [5] Wasserman, S. and Faust, K. (1994) Social network analysis: methods and applications, structural analysis in social sciences. Cambridge University Press, New York City, New York, U.S.A.
- [6] Chakrabarti, S. (2003) Mining the Web: Discovering knowledge from hypertext data."Morgan Kaufmann Publishers, USA.
- [7] Newman, MEJ. (2001) Co-authorship networks and patterns of scientific collaboration. In Proceedings of the National Academy of Sciences, 101(1), pp. 5200-5205
- [8] Davis, G.F. and Greve, H.R. (1997) Corporate elite networks and governance changes in the 1980s. The American Journal of Sociology, 103(1), pp. 1-37.
- [9] Stuart, T.E. (1998) Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. Administrative Science Quarterly, 43(3), pp. 668-98.
- [10] Chelmiss, C. and Prasanna, V.K. (2011). Social networking analysis: A state of the art and the effect of semantics. In Proceedings of 3rd IEEE Conference on Social Computing (SocialCom), Boston, MA, pp. 531-536.
- [11] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), pp. 440-442.
- [12] Milgram, S. (1967). The Small World Problem. Psychology Today, 2(1), pp. 60-67.
- [13] Ahn, J., Oh, D. and Lee, J. (2014) The scientific impact and partner selection in collaborative research at Korean universities. Scientometrics, 100 (1), pp. 173-188.
- [14] Patel, N. (1973) Collaboration in the professional growth of American sociology. Social Science Information, 6, pp. 77-92.
- [15] Wray, K. B. (2002) The epistemic significance of collaborative research. Philosophy of Science, 69, pp. 150-168.
- [16] Beaver, D. D. (2004) Does collaborative research have greater epistemic authority? Scientometrics, 60(3), pp. 399-408.
- [17] Guerrero-Bote, V. P., Olmeda-Gomez, C., and Moya-Anegon, F. (2013) Quantifying the benefits of international scientific collaboration. Journal of the American Society for Information Science and Technology, 64(2), pp. 392-404.
- [18] Zervas, P., Tsitmidelli, A., Sampson, D.G., Chen, N.-S., and Kinshuk. (2014). Studying research collaboration patterns via coauthorship analysis in the field of TeL: The case of Educational Technology & Society journal. Educational Technology & Society, 17 (4), pp. 1-16.

- [19] Newman, M. E. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2), pp. 404–409.
- [20] Farashbandi, F. Z., Geraei, E. and Siamaki, S. (2014). Study of co-authorship network of papers in the Journal of Research in Medical Sciences using social network analysis. Journal of Research in Medical Sciences, 19(1), pp. 41–46.



# **Surveying about Different Procedures of Intensifying Cloud Security and Presenting New Idea Based on Kalman Filtering**

Soudabeh Salehi Rekavandy  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

Maryam Davodpour  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

Mahdiye Abbassi  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

Golnaz Parvaneh Hosseini  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

Golnaz Shahideslami  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

Hooman Kashanian  
Islamic Azad University of  
Ferdows, Department of  
Computer Engineering,  
Ferdows, Iran

---

**Abstract:** Cloud is a collection encompassing a number of computers and servers that we find access to and is hosted by the internet. This technology provides capabilities for processing, storing and retrieving, support and etc. by way of the internet. Cloud computing consists of trustable services by future generation data centres that are based on virtual calculations and technologies and are provided by the internet. Now a day, numerous technologies are migrating from traditional systems to cloud. Off course, cloud can also be used as a tool for military and non-military goals. On the other hand, in this wide network where users are individuals and companies, in addition to shared use of resources, personal data is also placed in the network, need for security is very important. Frequently, users and organizations become more commonly and rapidly interested in using it when there is confidence in presence of high security methods in this service. In this paper, initially the effective role and advantages of cloud computing is evaluated in some large international military and non-military companies and organizations and ultimately security solution based on the Kalman filter is presented.

**Keywords:** Cloud Computing, Kalman Filter, Effect of Security in IT industry, Cloud Computing, Estimation and prediction

---

## 1. INTRODUCTION:

The background of computer science has persistently witnessed revolutions and fundamental changes. For example, it can be noted that in the first generation of computers that existed between the years 1945 and 1956, vacuum lamps were taken advantage of inside the computers. Later and with introduction of integrated circuits, they were used in third generation computers available in the years between 1963 and 1971 [1]. Ultimately, the fourth generation which begins from 1971 up until now alongside with development of technology large scale integration (LSI), very large scale integration (VLSI) and ultra large scale integration (ULSI) were used. Now a day also a technology named cloud computing is in the process of creating a new era in the computer industry and processing power. With study of historical trend, we can attest to the point that the idea of cloud computing sources from where processing powers can be placed in the hold of other users when current user or users do not need it [4]. The most common definition for cloud computing is access to power and processing even by way of conventional computers [7].

Examples of significant advantages of this technology are the possibility of sharing and concurrent work on a particular project or file as the following example. Assume that you want to process an individual or a special group of people's DNA [2]. If you are a cloud computing

user, you will be able to first use its extraordinary stronger processing power because your system might not have the necessary processing power. Second, you will be able to perform these very heavy operations concurrently with several of your friends and observe each other's work output on the project instantaneously [5]. Among its other potentials is that up to now you had to take advantage of your systems processing resources individually and once in a while you were forced to update your system hardware and software to be able to both access the most current software versions and also had need for hardware power for supporting your software. This by itself involves very high and in some cases unacceptable costs. In addition, if we wish to install this system on a higher number of systems, we need to purchase the license for use of the software in higher number of systems [9]. This reason contradicts the efficiency of cloud computing, because in cloud computing whenever we exit our account and reenter at another time, we witness that we will be using a newest version of the software without need to advance our systems hardware. The reason is that the necessary hardware power has been provided by cloud. In addition, you have not paid more fees for this up to dating. In traditional systems, if you have very confidential and important information on your system, constantly due to the worry of losing or theft or manipulation of your information, you will need to update your antivirus and other security

software (such as firewalls) [9]. If you would one day accidentally forget to perform this task, it would be possible that on the same day you would experience an attack, destruction or theft of your information. Yet, in cloud computing, you are constantly able at any time you enter your account to witness that your security system is equipped with updates made a second before. Among other advantages of cloud computing is knowledge of all suffixes. In other words, if you receive a file that is unknown in your system-since you did not know with what software to open it-with use of cloud computing, the file will be opened with the relevant software. In addition, if you own a company, you will be able to transfer your company's internal network onto the cloud computing and witness both increased speed and power and if like the past you use a server, you may not always have the amount of processing that can constantly keep the server busy and you just paid a high cost for electricity, repair and maintenance. These cases are just a corner of expansive efficiencies of cloud computing technology known as the next big thing [11].

## **2. STUDY OF THE DEEP EFFECTS OF CLOUD COMPUTING ON VARIOUS INDUSTRIES**

In the prior two sections, we referred to a selection of fundamental definitions and advantages of this newly evolved technology. In this section, we will note the most major applications of this technology and with finding the sensitivity of this technology, in most of its applications, modern security strategies will be suggested by authors of this paper [6].

The major reason for using cloud computing is that humanity in the present era tends to use technologies and wireless methods of communication. Meanwhile, Amazon is the first provider of commercial cloud computing. Another reason for this amount of use is that we prefer that the servers and level of processing and other hardware not be defined and to have an unbelievable power in processing and storing and other processing operations. The third

reason that can be mentioned for use of cloud computing instead of traditional systems is that now a day, humans gravitate to use of social sites. For example, the social site Facebook currently has more than 600 million registered users and everyday a huge number is added to this number. Another reason for its use is that the subscriber can decide which of its information to share with whom and in which places in the world while in traditional systems this possibility did not exist at all. Another reason is that in this method there is no longer need for separate kinds of programs, servers and systems for each group and company. Instead, each person with his or her own taste is able to use cloud computing. In addition, among other reasons we can mention lack of need for specialization in use, lack of need for a minimum level of hard and software in use and well known file formats in this system [8].

The Google Company is in the process of publishing the Chrome Factor System by way of which users can find themselves in cloud and to provide the capability of this kind of processing for them [4]. The Microsoft Company also after evaluation of this technology and the fear of losing its users is in the process of preparing Microsoft Azure which is a cloud factor system. The General Service Administration Company due to wide visits of its sites and the fear of destruction and hanging of its site decided to change hosting its site to cloud computing. National Aeronautics and Space Administration (NASA) with the NEBULA platform which is totally based on cloud computing has provided the possibility of people's participation in space projects and as a result wider possibility of storing and processing has been provided for NASA [10].

Department of Interior: this ministry which provides services for numerous federal organizations has decided to provide services based on cloud these days. Department of Health and Human Service: is in the process of proving platforms based on cloud computing for better and faster service provision. Census Bur: provider and director of SaaS services in the Salesforce site serves millions of people by way

of cloud [9]. The White House: is in the process of transferring its system to cloud computing technology so in this way it becomes able to have automatic voting system, direct conversation with citizens and have their internal network system working faster and with increased comprehensiveness [11].

Additionally, the government of England has operationalized the wide public network of G-Cloud on the basis of which they can have higher precision and speed in governmental affairs. Europeans also are in the process of taking advantage of this new technology in the following governmental and public sectors: management of public sector housing, transportation service networks, census, economic development, health services and contracting and education services. Denmark has also recently with two experimental plans Digitalise'r.dk and NemHandel tested cloud computing technology for its home users and in the opinion of the Danish regional government satisfaction with the results was because of cloud computing [10].

The Japanese also with a cloud named Kasumigaseki Cloud operationalized in the Tokyo governmental-industrial locale thought of expanding cloud for use in all governmental sectors and are leaders in this area. They also have named the region and technology as a “green” environment. Of course, the latter is also a major part of the digitalization of the Japanese project [6].

China: In China also the government particularly in the northern sections of the country with a project titled “The Yellow River Delta Cloud Computing Center” is not just in the process of economic evaluation and development, but is also thinking about a government based on cloud and wider efficiency. Even the government of China in the city of Wuxi has operationalized the factory for cloud services to become enabled in providing cloud processing resources for all companies and offices [4].

Thailand: In this country GITS (Governmental Information Technology System) has designed and built a private cloud for the governmental sectors of Thailand and it is planned that soon

using it various electronic services are provided for citizens and governmental sectors.

Vietnam: The IBM Company alongside with the government and universities of the country are developing and expanding the plan for cloud provision in a private and public form.

NewZealand: is in the process of expansive study on the best effects and uses possible of cloud computing [1].

Now, we will review the amazing effects of this technology on the situation of some smaller companies. A company such as the YouTube that in 2006 had 30 million pages added to its site daily should at present use cloud computing technology. The reason is that it is possible that the servers no longer have the ability for the necessary processing at time of search among all clips and films and the site of this company might face problems. Of course, if cloud computing method is not taken advantage of, servers with higher prices need to be purchased and installed which are very time consuming and costly [7].

The SmugMug Company which is a website for picture sharing for easier and increased service (such as picture conversion, retouching and etc) and also due to increased level of file volumes decided to use cloud computing. Additionally, Google provided the Google App Engine that enables you to implement your desired applications on Google substructures. In this way, you are able to share the documents you wish with specific people or in general at a global level. This service is totally free by the IBM company and Google is in the process of operationalizing a network on the basis of these new calculations for universities so via it, universities can perform their activities in a more focused and stronger way and if students of other universities have ideas or opinions about the active project, they can have a share in it. The first universities that connected to this network are University of Washington, Carnegie-Mellon University, Stanford University, The University of California at Berkeley and Maryland University [3].

The NASDAQ Company that has a wide stock and fund data base has decided to share and sell



them. Yet, their company was concerned with server use and level of their space occupation and therefore, it used the S3 service of Amazon. Additionally, smaller companies such as Nimbus and Eucalyptus are recently providing computational, storing and applied services with a fee [2].

The Sun Micro Systems company is in the process of operationalizing modern data centers for hosting cloud computing applications and users. This company in various locations globally has provided sites with powerful servers for support so potential defects can also be avoided [5].

Networks such as My experiment ([www.myexperiment.org](http://www.myexperiment.org)) and nanoHub ([www.nanohub.org](http://www.nanohub.org)) currently are in the process of migrating to cloud computing for sharing results of research and investigational works with people and direct connection with them.

Additionally, cloud computing provides banking services such that you can gain access to services with payment of a fee and will be able to perform tasks to acquire a payment from other users or improve a service and sell it and overall a buy and sale system will be accomplished [7].

Ultimately, the authors of this paper predict that in a future not very far, numerous companies and individuals will be occupied with provision of cloud computing services and technologies. Technicians and designers should also perform their role in the development and expansion of cloud computing correctly and provide services in accord with user need, simplify their access, provide more appropriate structure, higher capacity and capability and the possibility of adding and omitting resources [2].

### **3. NEW SECURITY STRATEGY SUGGESTED BY THE AUTHORS:**

In the first and second sections, the principles of cloud computing and some wide advantages were evaluated. In the third section, application of cloud computing in some of the major companies and sectors was evaluated and analyzed carefully. Now, it is time for authors of the paper to introduce the new security idea.

As you know, many companies face weakness in processing power because they do not have the ability to use very powerful servers and computers-for example due to economic reasons. Also, if they have such powerful systems, they will not have the kind of processing that can constantly engage these systems and they will bear an extra cost of repair and maintenance. Therefore, the only suggestions at present to them is that cloud computing is used. Yet, when we suggest to them to use cloud computing, we need to give companies and organizations particularly public and military institutes peace of mind when it comes to security. On this basis, the following idea is presented [3]:

The Kalman filter is an effective recursive filter that estimates the situation of a dynamic system from a sequence of defected and confounded evaluations. An example of the application of this filter is presentation of precise and up to date data persistently regarding place and speed of a hypothesized object only by having a sequence of observations from its location. Each of these observations comes along with a small error. This filter is used in a wide domain of engineering applications from radar to computer visualization and is an important topic in the theory of control and engineering of control systems. Now a day, broad composites of Kalman filters have expanded: From initial formulas-which are now called simple Kalman filters- to Smith's generalized filter-information filter- and a range of root of squares filters that have been designed by Bayerman, Torenton and many others.

Unlike other estimation filters such as the Wiener1 (that can only be implemented on static systems and cannot include effects of preliminary conditions), Kalman filters can be implemented on both static and non static processes and include the preliminary conditions of processes in estimation, prediction, filtering or best accidental control algorithms.

### **4. CONCLUSION:**

The authors suggest if we equip the cloud computing system with the Kalman filter, we

will be able to perform the necessary prediction of level of needed resources, number of online users in a particular hour and tracking an intrusive individual in the network easily. The reason is that exactly like rocket prediction in this section also with attention to the kind of movements of the hacker and his or her movement analysis; we will be able to easily track and expose him or her.

## 5. REFERENCES

1. *The NIST Definition of Cloud Computing*". National Institute of Standards and Technology. Retrieved 24 July 2011.
2. "What is Cloud Computing?". Amazon Web Services. 2013-03-19. Retrieved 2013-03-20.
3. *Oestreich, Ken, (2010-11-15). "Converged Infrastructure". CTO Forum. Thectoforum.com. Retrieved 2011-12-02.*
4. *Schmidt, Eric; Rosenberg, Jonathan (2014). How Google Works. Grand Central Publishing. p. 11. ISBN 978-1-4555-6059-2.*
5. *Rochwerger, B.; Breitgand, D.; Levy, E.; Galis, A.; Nagin, K.; Llorente, I. M.; Montero, R.; Wolfsthal, Y.; Elmroth, E.; Caceres, J.; Ben-Yehuda, M.; Emmerich, W.; Galan, F. "The Reservoir model and architecture for open federated cloud computing". IBM Journal of Research and Development 53 (4): 4:1–4:11.*
6. *Kyriazis, D; A Menychtas; G Kousiouris; K Oberle; T Voith; M Boniface; E Oliveros; T Cucinotta; S Berger (November 2010). "A Real-time Service Oriented Infrastructure". International Conference on Real-Time and Embedded Systems (RTES 2010) (Singapore).*
7. *Dario Bruneo, Salvatore Distefano, Francesco Longo, Antonio Puliafito, Marco Scarpa: Workload-Based Software Rejuvenation in Cloud Systems. IEEE Trans. Computers 62(6): 1072-1085 (2013)*
8. *Voorsluys, William; Broberg, James; Buyya, Rajkumar (February 2011). "Introduction to Cloud Computing". In R. Buyya, J. Broberg, A.Goscinski.*
9. *Alcaraz Calero, Jose M.; König, Benjamin; Kirschnick, Johannes (2012). "Cross-Layer*

*Monitoring in Cloud Computing". In Rashvand, Habib F.; Kavian, Yousef S.*  
*10. Amies, Alex; Sluiman, Harm; Tong, Qiang Guo; Liu, Guo Ning (July 2012). "Infrastructure as a Service Cloud Concepts".*  
*11. Metzler, Jim; Taylor, Steve. (2010-08-23) "Cloud computing: Reality vs. fiction," Network World.*

# Design and Implementing Novel Independent Real-Time Software Programmable DAQ System using Multipurpose MCU and Sigma Delta

Zohreh Vahedi  
Department of Computer  
,Ferdows Branch, Islamic  
Azad University  
Ferdows, Iran

Maryam  
KaivanluShahrestanaki  
Department of Computer  
,Ferdows Branch, Islamic  
Azad University  
Ferdows, Iran

.Hooman Kashanian  
Department of Computer  
,Ferdows Branch, Islamic  
Azad University  
Ferdows, Iran

**Abstract:** The crucial features of many demanding applications like industry and aerospace are data acquisition and telemetry. It is vital to observe and analyse the real time performance, in launch vehicle systems, so that designs can be certified and tuneable factors could be regulated to intensification the act and competence. At present used DAQ structures are of augmented size, weight and turn out to be exorbitant and power hungry. This article introduces a new mission-independent real time software programmable DAQ system using multipurpose MCU and sigma delta ADCs are planned, taking into account size, weight, cost and act without compromise on precision, firmness and drift act. Additional digital filtering steps are also added to progress the system act. This system is proficient for direct connections with diverse pressure and temperature sensors which interfaces 32 low frequency channel and two high frequency channels. The system planned operates in two modes; one is data acquisition mode and another is program mode. Operative power lessening methods and wireless interface protocol between diverse data acquisition modules is also affected upon as avenues for future work.

**Keywords:** Data acquisition system; Digital filter design; Versatile systems; wireless protocol; sigma delta

## 1. INTRODUCTION

A Data acquisition (DAQ) system in common involves of sensors, DAQ measurement hardware and a computer with programmable software. Compared to customary measurement system, PC based DAQ systems explicit the processing power, productivity display and connectivity abilities of industry standard computers providing a more powerful, flexible and cost effective measurement solutions. The sensors are the transducing elements, may want electrical excitation in the form of voltage or current, and are provided by the DAQ system. The DAQ systems and Telemetry systems have advanced substantially over the years and being used to collect real time data from sources in order to aid observing, analysis, and control facility [1, 2]. For this purpose the flight transmitted telemetry data is received and put in storage in the ground station. The data of the earlier mission forms a vital and important source for the analysis and design in later missions.

Current data achievement system considers only 16 low frequency channels and is controlled by an eight bit microcontroller of PIC18F6xxx family. Our necessity is to produce a well-organized system with reduced size and weight which meet the signal conditioning and data acquisition necessities of launch vehicle telemetry, without compromising on precision and determination. Adjusting the current system with the existing microcontroller will increase the extent and weight of the system and it possess peripheral and I/O confines too. This revision describes the design of a data acquisition system based on 16-bit PIC microcontroller

of PIC24EPxxx family and sigma delta Analogue to Digital Converter's (ADCs) of ADS1218 and ADS1255 [3, 4]. The planned system interfaces directly with sensors accepting their low level analogue signal as input in case of ADS1218 and high level analogue signal as input in case of ADS1255. ADCs related with each channel performs the required signal

conditioning. The gain and anti-alias filter cut-off frequency are set by the substances of control registers on the chip. The major factors of signal conditioning such as gain, input signal offset and anti-alias filter cut-off frequency are digitally controlled, thus rendering the system versatile and reconfigurable. There are two modes of performance. One is program mode and another is data acquisition mode. The program mode purposes include writing to ADC's configuration registers and reading from these registers to confirm the integrity of data. Other purposes such as reset, self-calibration, system gain and system-offset calibration are also applied in this mode. In data acquisition mode, the system obtains the digital data from the ADC and posts the proper data to Data Processing Unit (DPU). Use of one extra RS-485 link in the output for check out purpose, is also reviewed as another major feature of this system. Overview of Digital filtering phases using moving average concepts also improves the efficiency of the system.

The block diagram illustration of complete system with a brief explanation, and timing details are debated in section II. The software association including modes of operations and flow diagram are defined in section III. Performance details and consequences are discussed in section IV. Section V accomplishes the work and brings out an overview for additional enhancements in the system capabilities.

## 2. DESIGN DETAILS

### 2.1 Block Diagram Explanation

The hardware association of the system is shown in the fig.1. The planned system edges 32 –low frequency channels and two high frequency channels. Each channel contains of 24-bit sigma delta ADC of ADS1218 in case of low frequency channels and ADS1255 for high frequency channels [3, 4]. Whole system is controlled by a single 16-bit microcontroller

of PIC24EP512GU814 and these microcontrollers feature built-in flash memory for program storage, Random Access Memory (RAM) for data buffering and support of a diversity of standard edges such as Serial Peripheral Interface (SPI) and Universal Asynchronous Receiver Transmitter (UART). The output interface is electrically well-suited to industry standard RS-485 and makes use of three MAX3443 devices for redundant observing. The ADCs are arranged and read by the microcontroller through the SPI port and here the digital output from each channel is read and written on to the Data RAM inside the microcontroller. The data matching to each channel is then communicated to a downstream DPU over RS485. Communication over RS-485 bus is controlled through the UART port. Supporting circuits like Power-On- Reset (POR) and Voltage Reference Generators are also comprised to the hardware, to make it more efficient.

## 2. 2 Timing Information

In usual operational mode the microcontroller polls the channels sequentially for data readiness. The order of polling can be as per a format deposited in memory. The edge between microcontroller and ADC is by means of the standard Serial Peripheral Interface (SPI). The communications over the SPI port are synchronised via a Serial Clock (SCLK) of 1MHz initiate by the microcontroller [5, 6]. The ADS1218 will cater to all low frequency quantities up to 60Hz bandwidth and have a master clock rate of least 1MHz and maximum 5MHz. ADS1255 channels are used for high frequency quantities up to 1.06 KHz bandwidth and have a master clock rate of least 0.1MHz and maximum 8MHz. The communication done by the RS-485 bus follows the standard asynchronous communication protocol. When command from DPU arrives, the command is sent to the DAQ unit over RS-485 bus, which is arranged in half duplex multi drop bus format. Then each unit decodes the address and the one whose base address equals the command word sends back a reply word to the DPU. The communication over RS485 output edges are based on the precedence, interrupt driven mode controlled by the microcontroller unit (MCU). The MCU is arranged in external crystal oscillator mode with an operating frequency of 16MHz so that equivalent instruction rate of 8MIPS (Million Instructions per Second). The Universal asynchronous Receiver Transmitter (UART) provisions a baud rate of 2MBPS. The UART commands are of 11 bits which needs 0.5  $\mu$ s for each bit transmission. The signalling rate of 2MBPS supported by the transceiver will ensure that the output port can tolerate a combined throughput of 1 data samples in every 32 microseconds [7, 8].

## 3. SOFTWARE SPECIFICS

The Integrated Development Environment (IDE) is an atmosphere which integrates dissimilar simulation tools and compilers to provide a single window solution to development and debugging. MPLABV8.92 is an IDE delivered by the microchip,

### 3.1 Program Mode

Functions such as reset, self-calibration, system- offset and system-gain calibration are applied in this mode based on the

commands FE, F0, F3 and F4 correspondingly. The RS-485 interface itself can be utilized to load the formation data for each channel. The programming of ADC chip is an off-line operation and it is arranged in such a manner as to load this data automatically in power-up onto its arrangement registers. These ADCs have diverse gain values and is based on the internal reference voltage and differential input voltage of the ADC. Diverse gain settings of ADC are revealed in Table 1.

**Table 1. Gain sets of ADC**

| V ref | Differential input Voltage variety | Gain |
|-------|------------------------------------|------|
| 2.5V  | 0-2.5V                             | 1    |
| 2.5V  | 0-1.25V                            | 2    |
| 2.5V  | 0-0.625V                           | 4    |
| 2.5V  | 0-312.5mV                          | 8    |
| 2.5V  | 0-156.25mV                         | 16   |
| 2.5V  | 0-78.125mV                         | 32   |
| 2.5V  | 0-39.0625mV                        | 64   |
| 2.5V  | 0-19.53mV                          | 128  |
| 1.25V | 0-10mV                             | 128  |

### 3.2 Data Acquisition Type

In this approach, the ADCs are polled constantly. The 16-bit data is read from ADC and deposited in the data memory of microcontroller. It is conceivable through SPI routine and this process constant until the microcontroller receives a command from DPU. This command demand is controlled using interrupts and the interrupt service routine manages the posting of reply to DPU. It is conceivable through UART routine. After sending the reply, the microcontroller proceeds to ADC polling.

### 3.3 SPI routine

The communication between MCU and ADC is conceivable through SPI through five hand shaking signals, the Chip Select (CS), Data Ready (DRDY), SCLK (Serial Clock), Data Input (DI) and Data Output (DO) lines. The separate ReaDY (DRDY) signal corresponding to each chip is used to check the data legitimacy before reading the corresponding channel. The CS lines are separately supplied to each chip when equivalent channel is to be accessed. The data handover is synchronised with a SCLK of 1MHz. The microcontroller polls channel serially, selects the ADC if data is ready, issues a Read Data (RDATA) command and reads the data over the SPI port.

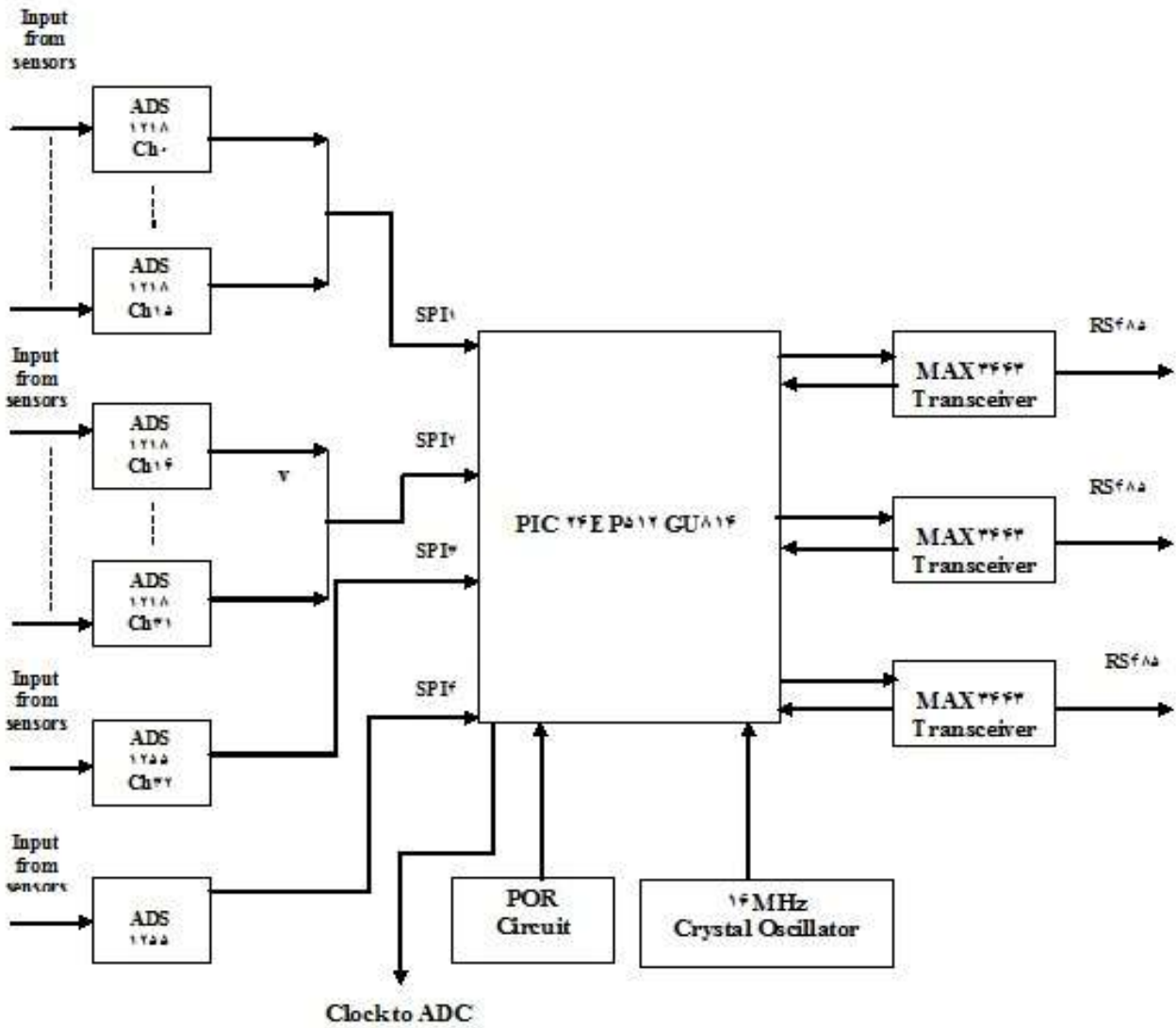


Figure.1 Block diagram of a 34-channel data acquirement scheme

### 3.4 Title UART routine

In the output, the microcontroller is interfaced to the DPU over UART. The UART uses the standard Non-Return –to-Zero (NRZ) format with one start bit, 8 data bits, 1 mode bit and one stop bit. The flowchart for UART routine is revealed in fig.2. The mode bit is used to distinguish between command and reply. The command order from DPU and the reply sequence from DAU (Data Acquisition Unit) in the RS-485 protocol are revealed in fig.3 and fig.4 correspondingly.

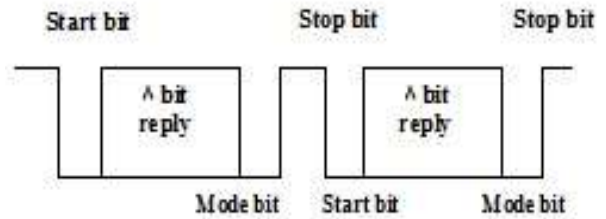


Figure.4 Reply order from DAU

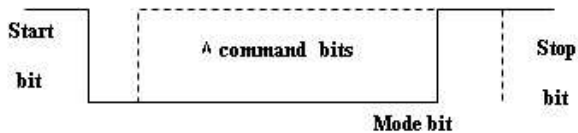


Figure.3 command order from DPU



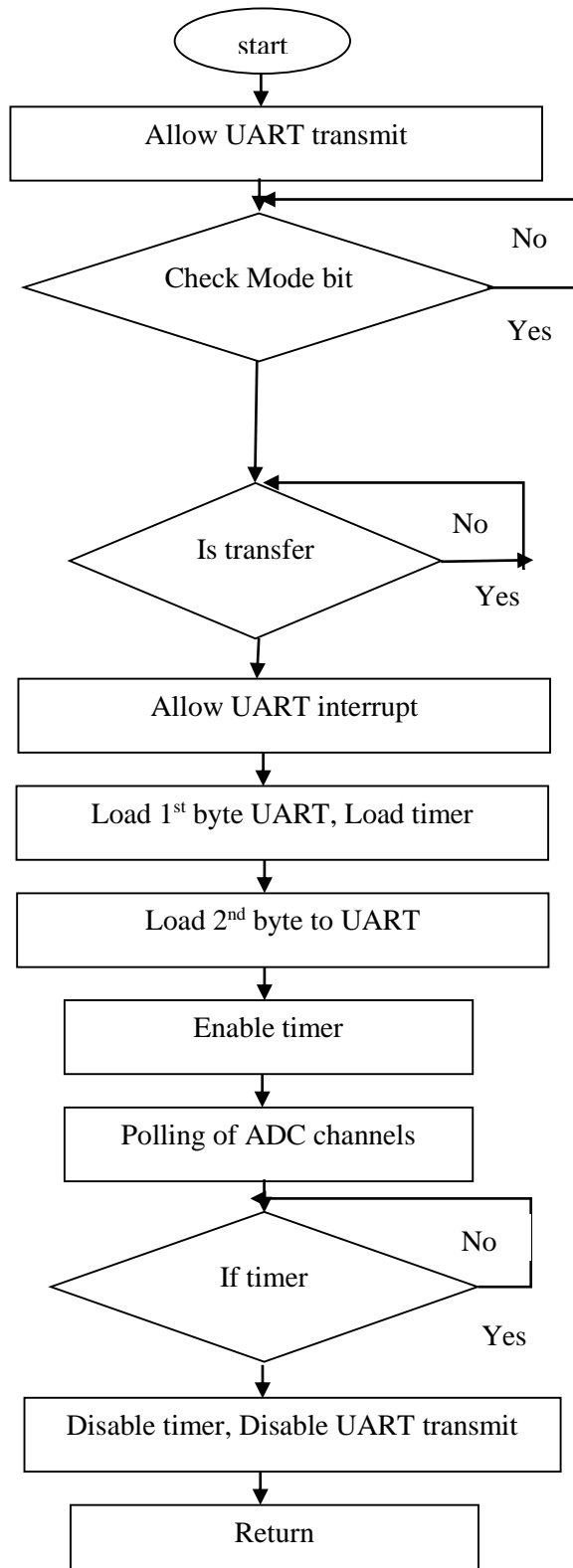


Figure .2 UART routine.

### 3.5 Filtering phase

A digital filter at the input of ADC using moving average notions increases the competence of the system. This simple low-pass filter creates an ADC with enhanced resolution and thus the system become more efficient. The Effective number

of bits (ENOB) in an ADC can be considered using the equation 1.

$$ENOB = \text{Log}_2 \left( \frac{V_{ref}}{3 * S \text{ standardDeviation}} \right) \quad (1)$$

- ENOB=Effective Number of Bits

### 4. CONCLUSION

In Data Acquisition method, the ADCs channels are polled constantly and the available data is read and stored in the memory of microcontroller. This system continuous until it receives a command from DPU. The command demand is managed using interrupts and interrupt service routine handles the posting of reply to DPU. The simulation consequences for SPI and UART routine in data acquisition method are shown in fig.5 and fig.6 correspondingly.

The program is verified in the demo board of PIC24E series. The board delivers a low-cost, modular expansion system for Microchip’s enhanced 16-bit Digital Signal Controllers (DSCs) or High-Performance Microcontrollers (MCUs). It also comprises of crystal oscillators, Green power indicator LED, USB connectivity for on-board debugger communications, three push button switches (SW1, SW3) for user-defined inputs, Three user-defined indicator LEDs (LED1, LED2, LED3), USB Type A connectivity for dsPIC33E/PIC24E USB host-based applications, Host mode power jumper and a Regulated +3.3V power supply for cause to move the starter kit via USB or an development board.

### 5. CONCLUSION AND UPCOMING WORKS

Data acquisition and telemetry is fragment of winning formula of many arenas comprising industry and aerospace. The carrying out of competent software programmable real time data acquisition system with limited size and weight is debated in this work. It outcomes in an efficient system in terms of sigma-delta ADCs, that can deliver higher input signal bandwidth and the

digital filter placed at the input of ADCs, which yields better-quality resolution. The adaptability of the system in terms of software reconfiguration, calibration and sensor interface extends its application.

Effort is on track of refining the system performance by considering power reduction concepts. Overview of a wireless protocol between diverse data acquisition module will also advance the performance [9].

| Address | 00   | 02   | 04   | 06   | 08   | 0A   | 0C   | 0E   |
|---------|------|------|------|------|------|------|------|------|
| 10F0    | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 |
| 1100    | FABE | FFE7 | F3F2 | FFFE | FFEB | FFEC | FFED | FFEA |
| 1110    | FEBA | FEAB | FEAC | FEAD | FEA7 | FEA8 | FEA9 | FE14 |
| 1120    | FE12 | FE1A | FE2B | FF3B | FF4C | FF5D | FF6A | FE7A |
| 1130    | FE8B | FE9C | FA53 | FA5B | FE17 | FE27 | FE47 | FE57 |
| 1140    | FE14 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 |
| 1150    | FFF3 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 | 0000 |

| Update | Address | Symbol Name | Value  |
|--------|---------|-------------|--------|
|        | 0224    | U1TXREG     | 0x0000 |
|        | 1002    | j           | 0x000B |
|        | 0100    | TMR1        | 0x0000 |
|        | 0248    | SPI1BUF     | 0x0001 |
|        | 0268    | SPI2BUF     | 0x0001 |
|        | 02A8    | SPI3BUF     | 0x0001 |
|        | 02C8    | SPI4BUF     | 0x0001 |

Fig.5 Polling procedure and reception of command by the microcontroller

| Update | Address | Symbol Name | Value  | Update | Address | Symbol Name | Value  |
|--------|---------|-------------|--------|--------|---------|-------------|--------|
|        | 0224    | U1TXREG     | 0x00AD |        | 0224    | U1TXREG     | 0x00FE |
|        | 1002    | j           | 0x0016 |        | 1002    | j           | 0x0016 |
|        | 0100    | TMR1        | 0x0000 |        | 0100    | TMR1        | 0xFFA7 |
|        | 0800    | IFS0        | 0x0800 |        | 0800    | IFS0        | 0x0800 |
|        | 0248    | SPI1BUF     | 0x0001 |        | 0248    | SPI1BUF     | 0x0001 |
|        | 0268    | SPI2BUF     | 0x0001 |        | 0268    | SPI2BUF     | 0x0001 |

Fig.6 Transmission of vital data requested by the DPU

## 6. REFERENCES

- [1] Arjun, R. ; Shahim, M.I.M. ; Kumar, G.S. July 2011 ,” Integrated environment for launch vehicle telemetry data management”, in Proc. IEEE .Int. Conf. Data Science & Engineering, Cochin, Kerala, pp. 170 – 174
- [2] Chandiramani, J.R. ; Bhandari, S. Jan. 2015; Hariprasad, S.A.,“Vehicle Data Acquisition and Telemetry”, in Proc. IEEE .Int. Conf. Signal and Image Processing, Jeju Island , pp. 186 – 191
- [3] Texas Instruments Incorporated; 2003–2013 , ADS1255 Data Sheet, USA .
- [4] Texas Instruments Incorporated ; 2003, ADS1218 Data Sheet, USA .
- [5] Microchip Technology Inc., 2005-2012,16-bit MCU and DSC Programmer’s Reference Manual
- [6] Amar Adane, Ania Ali Yahia, Essaid Mameri, Abd-El-Hamid Adane,August ,2015, ”Design of a Microcontroller-Based Data Acquisition System for Ground Weather Observations: Evaluation of Radio Refractivity of Air,” Int. J. Communications, Network and System Sciences(IJCNS), pp. 355-365
- [7] Microchip Technology Inc., 2009-2014 PIC24EP512GU814 Data Sheet.
- [8] Sreelal S., Smitha Jose, Preetha C. and Haresh Kumar Singh.2009 ,” A Compact Software Programmable Data Acquisition System”, IEEE Transactions.
- [9] Sabooj Ray, Jeba Arul Doss, Sheena Abraham, Pradeep N. and S. Prem Kumar, No. 3, March 2013, “Wireless Data Acquisition System for Launch Vehicles“,Defence Science Journal, Vol. 65, pp. 186-192.

# E-government Implementation in Kenya, an evaluation of Factors hindering or promoting e-government successful implementation

Francis Osanya Wamoto  
ICT officer  
The Judiciary  
Republic of Kenya

**Abstract:** Use of ICT in Government has ability to improve service delivery to its citizens, and yet many developing countries have lagged behind in the implementation of E-government. Many e-government initiatives also failed to achieve their objectives in developing countries. This paper therefore aims to identify critical failure or success factors in Kenya, using Heeks' Factor Model. A survey questionnaire was developed and data were collected and analyzed from officials and interested citizens. The analysis result enabled to highlight seven specific success and failure factors in Kenya. The enablers of e-government projects are good strategy formulation, internal and external drive, employee competence and enough funding, whereas main failure of e-government are weak ICT infrastructure, poor Project management and Design. Data analysis highlights both strengths and weaknesses for each factor or factor. Finally, the study ends up with recommendations for policy makers to shape the future of e-government system in developing countries.

**Keywords:** E-government implementation, Success and failure factors, Kenya.

## 1. INTRODUCTION

Today, e-government is well known to improve public service delivery. Its adoption results in greater efficiency and effectiveness of government administration. Therefore, e-government emerged as a hot issue in the academic circles and among policy makers from both the developing and industrialized nations [1]. In particular, the implementation of e-government initiatives has become one of the main goals in several countries including countries in the East Africa, more specifically the Republic of Kenya.

The Kenyan Government approved E-Government strategy in 2004, making the start of e-government journey. In 2006, National ICT policy was approved under the Ministry of Information and Communications (MoICT). The government has since initiated several e-government systems with the aim of enhancing efficiency, transparency and democracy within public administration. Some examples are e-Registry for business registration, passport application and processing system, and G-pay. "The overall goal of e-government is to make the Government more result oriented, efficient and citizen centered" [2].

However, e-government implementation is not a simple task. E-government initiatives have encountered different problems in its implementation and thus these have not been fully implemented in developing countries. The success or failure of e-government programs does not rely on whether they are implemented in developed or developing countries. Developing countries tend to have experienced more failures than advanced ones in implementing e-government [3]. Therefore, a thorough analysis on these problems is required to identify the challenges in developing countries.

This paper therefore aims to investigate critical failure or success factors that influence implementation of e-government systems in Kenya, using Heeks' Factor Model. A survey questionnaire was developed and data were collected

and analyzed from officials and interested citizens. Data analysis highlights both strengths and weaknesses for each factor or variable. Finally, the study ends up with recommendations for policy makers to shape the future of e-government system in developing countries.

## 2. ICT ENVIRONMENT SURROUNDING E-GOVERNMENT IN KENYA

After the adoption of the Vision 2030 roadmap, Kenya has largely experienced dynamism in the telecommunication sector hence national development plans have sought to take advantage in exploitation of its potential by promoting business process outsourcing (PBO) and IT-enabled services (ITES). Figure 1 illustrates the growth of mobile and internet penetration in Kenya since 1999 until 2012.

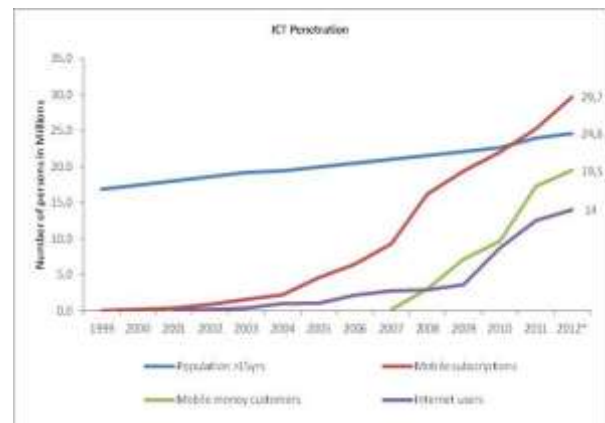


Figure 1. ICT penetration (Source: Julisha report (Kenya))

The liberalization of the telecommunications market started in the year 1999; this would then give a wide scope for the private sector innovation and market entry. At around the same period, the Communication Commission of Kenya (CCK) was established with a mandate of regulating the sector hence issued for the first time ISPs with licenses. Monopoly was experienced where Telkom Kenya Ltd operated the internet gateway and backbone until 2004. The national backbone was expanded during this period, however the international bandwidth never increased until the end of Telkom's exclusivity period.

The first e-Government strategy was formulated in 2004 and was approved in December, 2004. This was to create order and harmony in Government ICT initiatives which were at the time invariably characterized by disharmony and lack of coordination with each department pursuing their own ICT agenda which resulted to wastage through duplication of resources. Therefore, the key strategy was to set up ICT institutions that would immediately address this. Directorate of e-Government was to oversee, among others mandates, coordination of implementation of strategy, formulation and communication of ICT guidelines and enforcement of national and international standards (roles being played by ICT Authority after merging E-government Directorate, GITS and ICT Board)

The following are some of the projects that were implemented during the five year plan. Key among them is the Integrated Financial Management Information System (IFMIS) and Integrated Personnel and Payroll Database (IPPD) which are fully operational in the ministries. Other applications that have been rolled out include the Local Authorities Integrated Financial Operations Management Systems (LAIFOMS), Education Management Information System (EMIS), Integrated Taxation Management Systems (ITMS) currently known as ITAX after making great improvement in design, online Recruitment and Selection System in the public service commission and the Border control System in the Ministry of state for Immigration and Registration of persons.

The current state of Kenya e-Government system is for general administration including personnel, finance, procurement system, budget, email system among others. They are constructed as main online systems the public service provides for daily life to its citizens. Most systems run in individual departments for instance the national tax system, immigration information system, legal information system and education system. However, information exchange in these department is manual. The National ID, Passport tracking service, Examination result and candidate selection system and public tender system are also provided in the form of e-service though are limited as partial service [4].

### 3. SUCCESS OR FAILURE FACTORS OF E-GOVERNMENT IN KENYA

According to [2], e-Government is defined as "the use of ICT such as the wide area network, internet, and mobile computing, by government agencies to transform government operations in order to improve effectiveness, efficiency, service delivery and to promote democracy." Similarly, [5] defines e-government as the use of ICTs to improve public services delivery, and it brings with the promise of greater efficiency and effectiveness in the public sector."

Many studies have been carried out in order to identify critical success factors (CSFs) that influence e-government initiative implementation. [6] Identified 13 CSFs that are associated with e-government implementation. On the other hand, [7] formulated 27 CSFs just to mention a few. Ultimately, such success or failure factors that many researchers have suggested are well summarized into a Factor Model (FM) developed by [8] as an assessment framework for e-government in developing countries.

The FM highlights failure, success and improvisations areas as well as gap between design realities. It particularly expands the objectives and values dimension into elements on the drivers, strategy and change management [9]. The FM model comprises Drivers (external and internal pressure), Enablers (Strategy, management, design, competences, technology and others), and Constraints (considered as the negative aspects of drivers and enablers) as explained in Figure 2. This study chose the FM model because it focuses on human factor known to be critical to e-government projects successes and failures [10]. The FM model enables to identify and prioritize a list of e-government projects' success and failure factors based on perception of stakeholders in Kenya. This model summarizes the reasons behind success and failure of e-government projects. Left-pointing items encourage failure; right-pointing items encourage success.

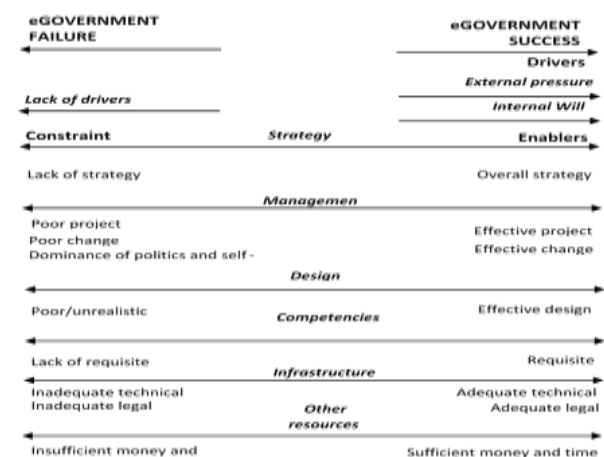


Figure 2. A factor model for e-government success and failure

This study therefore used Hicks' Factor Model to identify critical success factors (CSFs) that influence e-government initiative implementation in Kenya. First of all, internal and external drive for e-government is very important for successful implementation. External pressure relates to support for reform and achievement of e-government projects' goal from outside government such as aid donors, civil society and citizen, whereas internal pressure is associated with support from key internal government officials for reform and achievement of e-government projects' goals [8]. In order to achieve a successful project in civil agencies it has to be endorsed by top management [11]. Strong support from top officials is required throughout the implementation process in order to avoid e-government implementation facing obstacles such as resistance to change. Pressure from citizens for quicker services from government is also a good drive that can improve e-government implementation. There is a significant pressure from citizens in the Kenyan environment for faster services from government to citizens hence



pressurizing the government to improve service delivery by adopting and implementing e-government services. In the Kenyan context internal drive is highly indicating that there is a good political will towards e-government implementations. Another important factor for e-government implementation is strategy, and success in providing public services online highly depends on the strategy undertaken or adopted. Strategy relates to availability of overall vision and master plan for good governance and for e-government projects' implementation [8]. Kenya has taken a leading role in e-government implementation as compared to her other sister countries in East Africa. She has an approved ICT Master plan and an e-government strategic plan which is a roadmap to guide improvements to the online service to her people. Strategy works hand in hand with vision and therefore e-government implementation begins with establishing a broad vision hence forming a goal of any organization.

Management is related to three factors such as project management to deal with clear roles and planning, change management as incentives to create commitment, and change agents for projects. [12] Argues that change management in e-government implementation is a very complex issue. However, change management has not been enough considered because e-government is predominantly seen only as a technology mission, not organizational transformation issue. Changes expected are mostly caused by policies and legislation, public-private partnership and finally a huge resistance to change. Change management is a structured approach to transitioning individuals, teams and organizations from the current state to the desired future state [12].

Another much overlooked factor is design. Design relates to the availability of effective modular and incremental approach. Design requires stakeholders to participate at or involve in the inception of projects. Many e-government projects have failed to meet the needs of the users because stakeholders' involvement is overlooked in design development. In the Kenyan case, there might be more underlying factors that caused the election system to fail.

An innovation with substantial complexity requires more technical skills and needs greater implementation and operation efforts to increase their chances of adoption. [11] Insist that e-government adoption need a number of technical staff to help implement it. Many governments have suffered from lack of technical staff and are therefore very important component of factor in e-government implementation. In similar, competency factor was considerably weak and therefore classified as a constraint in e-government implementation in Kenya.

Infrastructure is a backbone of any e-government implementation initiative in both developed and developing countries. In many developing countries, infrastructure appears to be the main challenge why their e-government projects fail. The same scenario is not exceptional for Kenya. Infrastructure that is capable of supporting and enabling the execution of e-government is a requirement for successful e-government implementation [11]. An e-government infrastructure includes an infrastructure application server environment and its security, data and content management tools, application development tools, hardware and operating system, and systems management platform [11].

Last, funding is a critical factor for both starting e-government initiatives and continuing their implementation. Therefore, the

importance of funding in providing excellent service cannot be over emphasized. It is the glue that holds the building, collections and staff together and allows attaining goals [13]. They further insist that funding is essential in providing excellent service to citizens through excellent service delivery mechanisms.

#### **4. DEVELOPMENT OF STAKEHOLDER'S PERCEPTION QUESTIONNAIRE**

The questionnaire was prepared after a detailed study had been made of the previous literature and ICT environment in Kenya. The first part of the study focused on collecting the relevant documentation so that a more comprehensive picture of e-government system in Kenya could be determined in order to prepare survey questionnaires.

The possibility of surveying practices of e-government implementation by applying a limited number of questionnaires to the Ministry of ICT, ICT Authority, the IEBC and the Judiciary was therefore examined. This received good positive responses from many government officers and citizens. The survey was considered timely and valuable because Kenya had implemented e-government system since 2004 and no such study had been performed. To accomplish this, critical success and failure factors were constructed together with their relevant elements to measure the stakeholder's perception of e-government in Kenya.

Seven factors with a total of 33 variables were selected from the literature survey. Twenty three of these were based on Heeks' model (2003) which included drive, strategy, management, design, competencies, infrastructure, and other resources. The remainder were added to and deleted from each factor by reviewing a wider range of literature as well as Kenyan e-government system.

Based on this background, the questionnaire was constructed. The wording was intentionally directed towards all government officers and citizens who would respond to it. Following the pilot study the questionnaire was amended and some wording was changed. The final questionnaire was divided into two parts and the ratio scaling approach used. Part I concentrated on questions which would identify or categorize respondents; they included their organization, position/title, gender, etc. Anonymity was preserved. Part II was originally designed to measure 33 elements divided into seven critical factors of e-government systems using an interval response scale. Five point Likert scale questions were used and respondents were supposed to rate their responses. Stakeholders were able to indicate their perception of how well the e-government systems were run in the government (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). This means that 1 represents weakness with 3 being moderate and 5 being strength. Therefore, a score of 3 was regarded as the neutrality point to differentiate the success or failure factor. Finally, the stakeholder's overall perception of benefits from e-government implementation was sought.

#### **5. STATISTICAL ANALYSIS OF RESULTS SURVEYED**

The purpose of data collection was to test and assess the success and failure factors of e-government application implementation in Kenya. The Ministry of ICT, ICT

Authority, the IEBC and the Judiciary have a good number of ICT employees totaling to approximately 2000. The population of interest in this study comprised of the staff based in Nairobi and other major towns e.g. Mombasa, Kisumu Eldoret and Nakuru to represent the urban setting while the rest of the towns in Kenya to represent the rural setting. This study focused on the Government to Citizen Domain and therefore chose a small sample from the ordinary Citizens from five major towns of Nairobi, Mombasa, Kisumu Nakuru, and Eldoret (Urban-setting) and other towns as rural setting who were interested in e-government adoption. From these samples of the stated towns the study chose randomly thirty respondents from the general public. The data gathered was obtained using structured closed ended questionnaires for a period of three months, from 24th March to 21st June, 2015. The authors distributed 350 questionnaires to the ICT officers and other staff members in the Judiciary, the Ministry of ICT, the IEBC and ICT Authority, Kenya Revenue Authority, Ministry department Agencies and to the general public. 326 responses were received, however 17 responses were disqualified for various reasons such as lack of knowledge of any existing e-Government applications. The remaining 308 responses (88 % of total) were accepted as valid. According to [14], a response rate of 50% or more is ideal for data analysis. Table 1 provides a general overview of the Kenya officials who participated in this study in terms of demographic information including gender, age bracket, position they hold and their town of residence.

**Table 1. Demographic Information of Respondents**

| Demographic Information of Respondents |                 |           |      |
|--|-----------------|-----------|------|
|  | Variable        | Frequency | %    |
| Gender                                 | Female          | 124       | 40.3 |
|  | Male            | 184       | 59.7 |
|  | 18-24           | 11        | 3.6  |
|  | 25-34           | 140       | 45.5 |
|  | 35-44           | 107       | 34.7 |
| Age                                    | 45-54           | 48        | 15.6 |
|  | 55-64           | 2         | 0.6  |
|  | ICT Head        | 37        | 12   |
|  | ICT Staff       | 177       | 57.5 |
|  | Project Manager | 82        | 26.6 |
| Position                               | Other Positions | 12        | 3.9  |
|  | Nairobi         | 188       | 61   |
|  | Mombasa         | 10        | 3.2  |
|  | Kisumu          | 13        | 4.2  |
|  | Nakuru          | 14        | 4.5  |
| Town of Residence                      | Eldoret         | 32        | 10.4 |
|  | Other Towns     | 51        | 16.6 |

## 6. RELIABILITY AND VALIDITY VERIFICATION

The reliability of an instrument refers to the degree to which the instrument is free of random error. It is concerned primarily with consistency and stability of the measurement. Internal consistency tends to be a frequently used type of reliability in the IS domain [15]. Validity is often defined as the extent to which an instrument measures what it purports to measure and validity requires that an instrument is reliable, but an instrument can be reliable without being valid [16]. In this study Cronbach's coefficient alphas calculated based on the average item to item correlations was used to measure internal consistency. The results are presented as shown below in table 2. The values of the study instrument are reliable. The acceptable Cronbach's alpha value range from 0.7 to 0.95 and that means the cut-off point is at 0.7. [17] Says values below 0.7 can realistically, be expected because of the diversity of constructs being measured.

**Table 2. Cronbach's Alpha Reliability**

| Factor                    | No. of Items | Cronbach's Alpha | Mean   | Overall Rating |
|---------------------------|--------------|------------------|--------|----------------|
| Drive                     | 4            | 0.716            | 4.0048 | 7.1            |
| Strategy                  | 4            | 0.745            | 3.885  | 7.2            |
| Management                | 4            | 0.813            | 3.517  | 6.5            |
| Design                    | 5            | 0.765            | 3.431  | 6              |
| Competencies              | 4            | 0.69             | 3.364  | 5.8            |
| Infrastructure            | 5            | 0.756            | 2.647  | 4.6            |
| Other Resources (Funding) | 3            | 0.685            | 3.065  | 5.4            |

In order to verify whether all questions used in the questionnaire for measuring variables in the research model were suitable as measuring tools, the researcher carried out a factor analysis by AMOS. Confirmatory Factor Analysis by AMOS standard values are recommended that GFI, NFI and TLI be higher than 0.9. AGFI value to be more than 0.8 and RMR be lower than 0.05 [18]. The instrument used in this study met almost the required criteria required a part from Infrastructure factor. The result of this analysis is as shown in the Table 3 below

**Table 3. Confirmatory Factor Analysis**

| Construct | Initial items | Final items | GFI   | AGFI  | RMR   | NFI   | $\chi^2$ | P-value | Q      |
|-----------|---------------|-------------|-------|-------|-------|-------|----------|---------|--------|
| DR        | 4             | 4           | 0.978 | 0.889 | 0.039 | 0.960 | 13.783   | 0.001   | 6.891  |
| SA        | 4             | 4           | 0.985 | 0.926 | 0.029 | 0.968 | 8.880    | 0.012   | 4.440  |
| MA        | 6             | 4           | 0.994 | 0.970 | 0.017 | 0.991 | 3.642    | 0.162   | 1.821  |
| DA        | 5             | 4           | 0.978 | 0.889 | 0.039 | 0.960 | 13.783   | 0.001   | 6.891  |
| CA        | 4             | 4           | 0.967 | 0.835 | 0.075 | 0.902 | 21.657   | 0.000   | 10.829 |
| IA        | 5             | 3           | 1.000 | -     | 0.000 | 1.000 | -        | -       | -      |
| OA        | 4             | 4           | 0.994 | 0.971 | 0.037 | 0.986 | 3.512    | 0.173   | 1.756  |

### 6.1 Convergent Validity

Convergent validity is a function of the association between two different measurement scales which are supposed to measure the same concept, and is achieved when multiple indicators operate in a consistent manner [19]. [20], recommended that composite reliability (CR) should be equal to or greater than .60, and average variance extracted (AVE) should be equal to or greater than .50 [20]. The results of CR and AVE are presented in Table 4

**Table 4. Convergent Reliability**

| Item | Standardized estimate | S.E.  | AVE   | CR    |
|------|-----------------------|-------|-------|-------|
| DR1  | 0.702                 | 0.475 | 0.507 | 0.837 |
| DR2  | 0.791                 | 0.330 |       |       |
| DR3  | 0.670                 | 0.225 |       |       |
| DR4  | 0.680                 | 0.543 |       |       |
| SA1  | 0.590                 | 0.652 | 0.501 | 0.790 |
| SA2  | 0.670                 | 0.518 |       |       |
| SA3  | 0.750                 | 0.657 |       |       |
| SA4  | 0.803                 | 0.272 |       |       |
| MA1  | 0.730                 | 0.470 | 0.520 | 0.814 |
| MA2  | 0.691                 | 0.466 |       |       |
| MA3  | 0.711                 | 0.466 |       |       |
| MA6  | 0.752                 | 0.500 |       |       |
| DA1  | 0.724                 | 0.442 | 0.532 | 0.808 |
| DA2  | 0.650                 | 0.408 |       |       |
| DA3  | 0.797                 | 0.351 |       |       |
| DA5  | 0.740                 | 0.811 |       |       |
| CA1  | 0.665                 | 1.140 | 0.506 | 0.705 |
| CA2  | 0.671                 | 0.774 |       |       |
| CA3  | 0.648                 | 0.940 |       |       |
| CA4  | 0.843                 | 0.378 |       |       |
| IA1  | 0.859                 | 0.373 | 0.592 | 0.755 |
| IA2  | 0.922                 | 0.196 |       |       |
| IA5  | 0.432                 | 1.020 |       |       |
| OA1  | 0.841                 | 0.420 |       |       |
| OA2  | 0.816                 | 0.389 | 0.502 | 0.709 |
| OA3  | 0.585                 | 1.267 |       |       |
| OA4  | 0.541                 | 1.102 |       |       |

## 7. DESCRIPTIVE STATISTICAL PERSPECTIVE

### 7.1 Gender of Respondents

Male respondents were slightly higher (59.7%) than the female respondents (40.3%) implying a normal ratio distribution in any setting. On a gender grouped T-test, there is a significant difference in Drive and Other (Funding) factors (see Table 5.5). The result revealed that male had high score than female respondents. All other factors showed insignificant levels, which means no difference male and female. The results are as shown in Table 5 below.

**Table 5. Grouped T-test on Gender**

| GENDER | Mean   | Std. Deviation | F      | Sig.   | t    | Sig. (2-tailed) | mean difference |         |
|--------|--------|----------------|--------|--------|------|-----------------|-----------------|---------|
| DR     | Male   | 4.1835         | .58183 | 1.148  | .285 | 3.252           | .001            | .22895  |
|        | Female | 3.9565         | .61305 |        |      |                 |                 |         |
| SA     | Male   | 3.9758         | .67732 | 1.595  | .208 | 1.852           | .065            | .15244  |
|        | Female | 3.8234         | .72855 |        |      |                 |                 |         |
| MA     | Male   | 3.6069         | .68534 | 12.508 | .000 | 1.620           | .106            | .15033  |
|        | Female | 3.4565         | .86651 |        |      |                 |                 |         |
| DA     | Male   | 3.4819         | .68534 | 3.096  | .079 | .998            | .319            | .08512  |
|        | Female | 3.3967         | .76479 |        |      |                 |                 |         |
| CA     | Male   | 3.2964         | .91140 | 2.240  | .135 | -1.130          | .259            | -1.1260 |
|        | Female | 3.4090         | .81913 |        |      |                 |                 |         |
| IA     | Male   | 2.7204         | .97705 | .573   | .450 | 1.122           | .263            | .12260  |
|        | Female | 2.5978         | .91541 |        |      |                 |                 |         |
| OA     | Male   | 3.2500         | .65176 | 3.065  | .081 | 3.618           | .000            | .30303  |
|        | Female | 3.0340         | .78354 |        |      |                 |                 |         |

### 7.2 Position of Respondents

A majority of the respondents i.e. 177 (57.5%) indicated that they work as ICT staff, followed by 82 (26.6%) who were Project Managers, 37 (12%) ICT Heads, and the least 12 (3.9%) were ordinary citizen with interest in e-government implementations in Kenya. The researcher carried a grouped T-test to find out the perception to all the factors between Managers and Operators. Managers included top ranking officials (ICT Heads) in the government whereas operators included ICT staffs and Project Managers. The results as shown in Table 6 below are significant more especially in

Drive, Strategy, Design, Competence and Infrastructure factors. Operator group were higher than managers in five factors, meaning that manager group have lower perception of these factors being driving force for successful implementation of e-government in Kenya. On the other hand, both management and funding factors showed no difference between operator and managers.

**Table 6. Grouped T-Test on Position**

| Position | Mean     | Std. Deviation | F      | Sig.  | t    | Sig. (2-tailed) | mean difference |         |
|----------|----------|----------------|--------|-------|------|-----------------|-----------------|---------|
| DR       | Manager  | 4.0116         | .62712 | 2.588 | .109 | -2.420          | .016            | -.22821 |
|          | Operator | 4.2398         | .47037 |       |      |                 |                 |         |
| SA       | Manager  | 3.8359         | .72811 | 3.158 | .077 | -2.801          | .005            | -.30695 |
|          | Operator | 4.1429         | .55199 |       |      |                 |                 |         |
| MA       | Manager  | 3.5010         | .79743 | .726  | .395 | -0.810          | .419            | -.10108 |
|          | Operator | 3.6020         | .82114 |       |      |                 |                 |         |
| DA       | Manager  | 3.3890         | .73300 | .051  | .821 | -2.326          | .021            | -.26407 |
|          | Operator | 3.6531         | .70492 |       |      |                 |                 |         |
| CA       | Manager  | 3.3127         | .86612 | 1.682 | .196 | -2.413          | .016            | -.31991 |
|          | Operator | 3.6327         | .76564 |       |      |                 |                 |         |
| IA       | Manager  | 2.5714         | .92269 | .010  | .919 | -3.300          | .001            | -.47619 |
|          | Operator | 3.0476         | .94526 |       |      |                 |                 |         |
| OA       | Manager  | 3.0299         | .72958 | .001  | .974 | -1.931          | .054            | -.22008 |
|          | Operator | 3.2500         | .74302 |       |      |                 |                 |         |

### 7.3 Residence of Respondents

A majority of the respondents 188(61%) reside in Nairobi, 32 (10.4%) reside in Eldoret, 14 (4.5%) in Nakuru, 13 (4.2%) in Kisumu, 10 (3.2%) in Mombasa and the rest 51 (16.6%) reside in other towns across the country. This shows that the survey was a representative of the entire country covering both urban and rural settings. On grouped T-test, there was significant difference between Cities (Nairobi, Mombasa, and Kisumu) and other towns (Nakuru, Eldoret and other small towns) in Design Assessment. Other factors did not post significant results. Respondents living in cities have high perception of Design factor as a success factor in implementing e-government than respondents living in other towns.

**Table 7. Grouped T-test on Respondents residence**

|    | N           | Mean | Std. Deviation | Std. Error Mean | F      | Sig.   | t    | Sig. (2-tailed) | Mean Difference |         |
|----|-------------|------|----------------|-----------------|--------|--------|------|-----------------|-----------------|---------|
| DR | other towns | 97   | 4.1134         | .40023          | .04064 | 11.274 | .001 | 1.279           | .202            | .09563  |
|    | cities      | 211  | 4.0178         | .68399          | .04709 |        |      |                 |                 |         |
| SA | other towns | 97   | 3.8119         | .62503          | .06346 | 3.167  | .076 | -               | .223            | -.10639 |
|    | cities      | 211  | 3.9182         | .74651          | .05139 |        |      | 1.220           |                 |         |
| MA | other towns | 97   | 3.6366         | .69599          | .07067 | 5.029  | .026 | 1.783           | .076            | .17451  |
|    | cities      | 211  | 3.4621         | .84041          | .05786 |        |      |                 |                 |         |
| DA | other towns | 97   | 3.6108         | .67118          | .06815 | 1.958  | .163 | 2.952           | .003            | .26248  |
|    | cities      | 211  | 3.3483         | .74808          | .05150 |        |      |                 |                 |         |
| CA | other towns | 97   | 3.3840         | .85396          | .08671 | .261   | .609 | .282            | .778            | .02976  |
|    | cities      | 211  | 3.3543         | .86143          | .05930 |        |      |                 |                 |         |
| IA | other towns | 97   | 2.6976         | .95268          | .09673 | .004   | .950 | .637            | .525            | .07358  |
|    | cities      | 211  | 2.6240         | .93705          | .06451 |        |      |                 |                 |         |
| OA | other towns | 97   | 3.1443         | .66618          | .06764 | 2.481  | .116 | 1.287           | .199            | .11589  |
|    | cities      | 211  | 3.0284         | .76323          | .05254 |        |      |                 |                 |         |



### 7.4 Age of Respondents

More so, a bigger number of the respondents, 140 (45.5%) were aged between 25 yrs. and 34 yrs., followed by 107 (34.7%) respondents aged between 35 yrs. and 44 yrs., the 48 (15.6%) respondents between 45 yrs. and 54 yrs., then 11 (3.6%) of respondents were between 18 yrs. and 24 yrs. and the least were 2 (0.1%) respondents who were aged between 55 yrs. and 64 yrs. The youths (age 18-44 years) have high perception about Strategy factor in enabling successful implementation of e-government than the old (45 years and above). There was a significant difference in Strategy factor as shown in figure 8 below.

**Table 8. Grouped T-test on Age**

|    |       | N   | Mean   | Std. Deviation | F      | Sig. | t      | Sig. (2-tailed) | Mean Difference |
|----|-------|-----|--------|----------------|--------|------|--------|-----------------|-----------------|
| DR | old   | 50  | 4.0300 | .63011         | .257   | .612 | -.226  | .821            | -.02136         |
|    | young | 258 | 4.0514 | .60721         |        |      |        |                 |                 |
| SA | old   | 50  | 4.0950 | .71409         | .064   | .801 | 2.300  | .022            | 25101           |
|    | young | 258 | 3.8440 | .70479         |        |      |        |                 |                 |
| MA | old   | 50  | 3.6000 | .95164         | 7.575  | .006 | .800   | .424            | .09903          |
|    | young | 258 | 3.5010 | .76921         |        |      |        |                 |                 |
| DA | old   | 50  | 3.4200 | .91116         | 11.938 | .001 | -.116  | .908            | -.01314         |
|    | young | 258 | 3.4331 | .69646         |        |      |        |                 |                 |
| CA | old   | 50  | 3.1950 | .98624         | 1.302  | .255 | -1.522 | .129            | -.20132         |
|    | young | 258 | 3.3963 | .82881         |        |      |        |                 |                 |
| IA | old   | 50  | 2.6333 | 1.05032        | 1.025  | .312 | -.114  | .910            | -.01654         |
|    | young | 258 | 2.6499 | .92062         |        |      |        |                 |                 |
| OA | old   | 50  | 3.0000 | .94761         | 12.257 | .001 | -.682  | .496            | -.07752         |
|    | young | 258 | 3.0775 | .68779         |        |      |        |                 |                 |

### 8. CURRENT STATUS OF E-GOVERNMENT SYSTEM IMPLEMENTATION

With regard to the status of e-government implementation, 194 (63% of total) respondents indicated that the e-government systems have been successfully implemented and 101 (32.8% of total) currently implemented. Meanwhile, 7 (2.3% of total) indicated that they are in the planning stage, whereas only 6 (1.9%) being failed.

IFMIS is widely known by the respondents (45.5% of total) with ITAX 16.2%. This means that IFMIS system is widely implemented across the government Offices and that is the most successfully implemented E-application in Kenya. The biggest responsibility for implementing the e-government lies in government (65.6% of total) with the second in public private partnership (26% of total). Minimal portion of responsibility (8.4% of total) is accounted for by International Development Agencies (Donors) and private agencies and the rest.

On the other hand, the findings illustrated in Table 4 reveals that 95.5% of the respondents strongly understand the benefits of e-government to citizens, whereas only 4.5% of the respondents disagree with the majority's view. The overwhelming support of e-government at 97.4% is a strong indication that Kenyans have embraced e-government, which will be a push factor for successful e-government implementation by the Government.

**Table 9. Support for E-government Implementation**

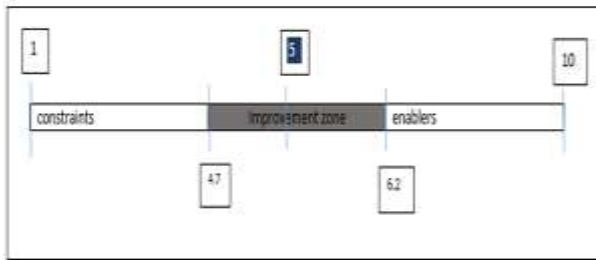
| Support for E-government Implementations in Kenya |               |           |      |
|---|---------------|-----------|------|
|   | Variable      | Frequency | %    |
| Individual Support                                | Support       | 300       | 97.4 |
|   | Don't Support | 8         | 2.6  |
| Understanding benefits of E-gov't to Citizens     | High Benefits | 294       | 95.5 |
|   | Low Benefits  | 14        | 4.5  |

### 9. ENABLERS AND CONSTRAINTS FACTORS IN E-GOVERNMENT IMPLEMENTATION

The stakeholder's overall perception was determined by combining all the constituent variables. This produced an average score of 6.086 as the overall perception of e-government implementation in Kenya. The illustrated graph in figure 4 highlights the observation based on the results of grouped responses of the level of e-government implementation strategy, showing areas of strength and weaknesses calculated by assigning one to strongly disagree (weakness, e-government failure) and ten to strongly agree. Having discussed this with several numbers of global ICT experts, the authors developed a model of evaluating the level of e-Government implementation. It was felt that a mean score above 6.0 was desirable, one below 4.7 indicated immediate corrective action to be implemented and that between 4.7 and 6.0 required gradual action for improvement of e-government implementation. In other words, the model includes the Zone of Constraint (less than 4.7 of total 10; Failure) the Zone of Improvement (between 4.7 and 6.2 of total 10; Partially Success) and the Zone of Enabler (greater than 6.0 of total 10; Success). From this model, The Kenyan e-government implementation belongs to the Zone of Improvement (6.0 of total 10). Figure 4 below illustrates how Zones were classified.



**Figure 3. Weighted value of individual factors**



**Figure 4. Zone Classification**

From the graph in figure 3, the area of weakness is in Infrastructure factor hence form a constraint. On the other hand, Strategy factor presented the strongest score in terms of e-government implementation, followed closely by Drive assessment and Management. They fall under the Zone of Enabler. The remainder Design, Other resources and Competency factor fall under enablers of Improvement. This means that the Kenyan government should focus her efforts to improve Infrastructure and then gradually improve Design, Competency and Other resources factors. In this section the study discusses factor by factor and identifies the weak areas that require improvement for successful implementation of e-government in Kenya with respect to their weighted score.

With regard to Drive factor in table 5, only three variables were tested and it's shown that all components scored considerably well. It reflects that there is strong drive both from within and outside government for e-government implementations in Kenya. The research also highlights that key government officials recognize e-government issues and have strong drive for their implementation, the two variables scored with almost similar margin. However, this should not be overlooked instead more effort has to be added in order to achieve 100 % implementation.

Strategy factor performed well in this study. All the components score fall under the Zone of Enabler. Existence of National ICT Master plan scored 7.8, followed by Existence of an Approved National ICT policy with 7.4, existence of an e-Government Strategic plan with 7.1 and participation of stakeholders in strategy development with 6.7 score. However, some respondents had reservation about participation of stakeholders in strategy development, which means that more effort for improvement is required in this component.

Six components were tested under management factor; Existence of a Champion organization in Kenya scored considerably high with 7.2, followed by resistance from workers or employees scored 6.9, perceived infighting among Government departments with 6.5, Strong consideration of risk in e-government implementation with 6.1 and lastly, Consistence and coordination of e-government projects scored 5.4. This study has identified two underlying components that hinder successful implementation of e-government projects. They are resistance from employees and infighting among government departments. 72.4% of the respondents concur that there is resistance to change from workers during e-government projects implementation in Kenya. There existed infighting among Champion organizations i.e. ICT Authority, MoICT and the Treasure due to duplicate mandates among them. As a clear evidence, 62.3% of respondent agree that there is infighting among Government Departments. The reasons behind this could be due poor project management and lack of clear roles and mandates among government

ministries and departments responsible for e-government implementation.

All components in Design factor scored between the Zones of Improvement and Enabler, presence of quick and feasible objectives scored highest of all with 6.7, followed by other components as shown in the table 9. Therefore, more efforts are required to improve design of projects in order to meet the needs of Kenyans as well as achieve the well-intended objectives of e-government applications.

The overall rating of Competence factor was 5.8. Many respondents who were interviewed possessed basic computer skills. However, it seems the government has not taken an initiative to sensitize its employees on e-government.

The infrastructure factor performed poorly and only managed to raise an overall rating of 4.6. The result indicates that there is no adequate Laws on e-transactions, little presence of Security and authentication technology, minimal presence of hardware, software and network technologies as well as minimal interconnections of government offices. A lot of improvement efforts are required for improvement because infrastructure is the bedrock of e-government that cannot be ignored under all circumstances.

Another poorly performed factor is funding of e-government projects. This factor scored an overall rating of 5.4 which falls under the Zone of Improvement. There is an indication that most e-government projects are funded by donors and private sector and this puts the government in a vulnerable position with regard to control of e-government initiatives. This study has therefore recommended among other things, that the government should take charge of e-government initiative as is the wish of many respondents and donors to play an oversight role.

### 9.1 Risk Assessment

All the factor rating score are added up and interpreted according to the table below. In this research, factors scored as follows; Drive Factor 7.1, Strategy Factor 7.2, Management Factor 6.5, Design Factor 6, Competence Factor 5.8, Infrastructure Factor 4.6 and Other Resources (Funding) 5.4. Making an Overall Score of 42.6, thus e-government implementation in Kenya might be a partial failure unless action is taken. As much there seems to be a slight improvement with comparison to the study by [1] where overall performance of e-government in Kenya scored 38.2, the score under this study still falls under the same category of partial failure. See table 10 below.



**Table 10: Risk Assessment layout**

| <i>Overall rating</i> | <i>Likely outcome</i>  |
|-----------------------|--|
| 0–13                  | Your e-government project will almost certainly fail unless action is taken.                     |
| 14–27                 | Your e-government project may well fail unless action is taken.                                  |
| 28–41                 | Your e-government might fail totally, or might well be a partial failure unless action is taken. |
| 42–55                 | Your e-government project might be a partial failure unless action is taken.                     |
| 56–70                 | Your e-government project may well succeed.  |

## 10. DISCUSSIONS AND RECOMMENDATIONS

E-government was established for the first time in Kenya in 2004 after the adoption of the first ever e-government strategy. So over the last ten years, the government of Kenya initiated major capital investments such as installation of ICT infrastructure. Development partners formed the biggest percentage of funding and the government contributions usually have been in form of technical and support staff and facilities that included buildings. So far the Government Information Technology investment and management framework has connected all Ministries to the internet under the executive Network [21]. The government has also connected all Ministries to run the Integrated Information Systems (IFMIS) and more others.

However, successful implementation of e-government initiatives requires careful planning and other issues that take into consideration a number of factors alongside technology. Many projects have failed because they took a technology-centric approach where technology has been seen as the main drive for e-government development. This study is therefore worthy because it describes the specific success and failure factors and their relevant elements in a particular country for assessing the stakeholder's perception of their e-government system. It is found that good strategy formulation, internal and external drive, employee competence and enough funding are enablers of e-government projects in Kenya as well as in any developing country especially in Africa because the underlying conditions and factors are quite similar.

The study too identified three factors which include poor ICT infrastructure, poor Project management and Design as main failure of e-government in Kenya. This however should not be tied only onto the mentioned seven factors, these were only identified as critical factors for e-government projects implementation in developing countries. Therefore among other things, the findings of this study highlight the need for further research focusing on more factors other than the ones evaluated in this study.

The research found out that among the main factors, Infrastructure scored very low below the minimum to form a constraint in Kenyan e-government implementation process.

Competence and funding (Other Resources) did not do well either. In order to address the constraints and improve on competence and funding factors, this study provides a variety of recommendations. First of all, Kenyan government should create clear role and mandate for champion organization. This study found out that there is significant infighting among government Ministries and departments and this could be due to unclear role definition and duplication, it is therefore important for Kenya to give greater mandate to one organization for instance ICT Authority for implementations of e-government projects as it was before when the Directorate of E-government was in existence.

Second, the Government should develop project planning and management capacities within government, to enable government to give some direction to, and have some control over e-government projects. There needs to be a continuous focus on raising the awareness, understanding and skills of those involved in e-government projects. This must include promotion of the benefits of the projects. In the Kenyan context, the study found out that employees are not well sensitized on e-government which should not be the case because these employees are the super users of e-government system hence they need to be sensitized as early as possible.

Third, the government focus on building telecommunication infrastructure and choose appropriate technology to its environment. This can be based on or building a modern telecommunication infrastructure.

Last, the government should avoid Big Bang approach of project implementation and instead introduce them gradually, systematic way in order to avoid resistance among users of the new technology or application. There should be initial goals and pilot-test the hardware, the database and the new procedures before any attempt to build up to full-scale implementation. E-government projects can be left to drift unless they are well managed. Therefore, e-government projects must include continuous monitoring, and continuous action to improve the project on the basis of that monitoring.

### 10.1 Further Recommendations

- **Create clear role and mandate for a champion organization.** This study found out that there is significant infighting among government Ministries and departments and this could be due to unclear role definition and duplication, it is therefore important for Kenya to give greater mandate to one organization for instance ICT Authority for implementations of e-government projects as it was before when the Directorate of E-government was in existence.
- **Encourage transparency.** The government should do whatever can be done to introduce transparency and rationality into procurement and recruitment processes. For example, it may be possible to promote use of a project Web site that gives details of project activities. Involvement of donors may introduce oversight procedures.
- **Training/awareness-raising.** There needs to be a continuous focus on raising the awareness, understanding and skills of those involved in e-government projects. This must include promotion of the benefits of the projects. In the Kenyan context, the study found out that employees are not

well sensitized on e-government which should not be the case because these employees are the super users of e-government system hence need to be sensitized as early as possible.

- **Dealing with politics.** Some way must be found to deal with the politics, sentiment and self-interest that inevitably forms part of e-government projects. The study therefore recommend that the president should take leadership and create a committee of experts who will directly report to him without any political interference.
- **Find political will.** The success or failure of e-government projects - especially those involving citizens and the democratic process - is significantly determined by the political context. Unless there is a political will to see the e-government project succeed, then it is likely to fail.
- **Involve civil society.** In projects that touch the broader issues of governance, the participation of civil society organizations should be encouraged from the inception. They can have longer-term objectives for improvement of the governance process that can usefully balance objectives in government that may be merely for short-term survival.
- **Establish partnerships.** Particularly with others who have undertaken similar initiatives.
- **Customize e-government goals to current realities.** Understanding the motivations and interests of key stakeholders will help designers of e-government systems fit with what is possible, rather than trying to design something that will not happen.
- **Customize timescales to current realities.** Ministers should not expect automatic and overnight transition to electronic networking. They must recognize that time and groundwork are needed.
- **Customize training to user needs.** Customize training content to the specific job needs of staff. This may well mean some initial study of work content to understand the role of information and communication (and, hence, the potential role of ICTs) in the work of government staff. In this case,
- **Get citizen representation.** Apart from experts and government officials, it is desirable to involve the representatives of different categories of citizens in project monitoring and implementation. They can be associated with high-level decision making and review committees.
- **Focus on affordable, feasible solutions.** The information needs of the institutions involved should be thoroughly assessed. During this process it is in most cases obvious that there is a big difference between those things which are truly necessary and those which people want having seen somewhere else, especially on the Internet. Solutions based on the latter may not be affordable on the long run. Sustainability of e-government initiatives that begun with the help of foreign donors

depends in the future on costs which the institution will have to pay out of its own budget. Sustainability will only be achieved by focusing on true and affordable needs, not on 'bells and whistles'.

- **Train users early via prototypes.** Early training efforts using rapidly-prototyped versions can make weaknesses of the system obvious at a stage where changes and modifications take less time than after making a system final and then handing it over to end users.
- **Focus on immediate benefits and high visibility with broad citizen benefit.** For an e-government project to succeed, the underlying technology need not be very sophisticated. What is required is an understanding of citizen needs, proper structuring of the project and early delivery of benefits. It is also important to have projects that have a fairly high degree of public visibility and which serve a broad spectrum of society so that people at large are convinced about the use of ICTs in government.

## 11. CONCLUSION

This study has shown that good strategy formulation, internal and external drive and Management are enablers of e-government projects implementation in Kenya as well as in any developing country especially in Africa because the underlying conditions and factors are quite similar. The study too identified four factors which include poor ICT infrastructure, poor funding, Poor Computer Skills (Competence) and Design as main failure causes of e-government in Kenya. This however should not be tied only onto the mentioned seven factors, these were only identified as critical factors for e-government projects implementation in developing countries. Therefore among other things, the findings of this study highlights the need for further research focusing on more factors other than the ones evaluated in this study. This study indicated a partial failure of e-government implementation and which require an urgent addressing of the underlying failure factors.

The outcome of this study can provide a big implication on the policy of any developing countries. It is felt that the government must take a clear and leading role in promoting e-government initiatives for success to be seen. The study has therefore provided several specific recommendations for Governments and policy makers in order to reduce the risk of failure of e-government projects. The reason being, developing countries may have similar characteristics to the Kenyan context which presents various challenges that affect the successful implementation of e-government. The reason is that developing countries may have similar characteristics to the Kenyan context which presents various challenges that affect the successful implementation of e-government.

## 12. REFERENCES:

- [1] N. Y. Awiti, 2012. [Online]. Available: [http://library.kaist.ac.kr/thesis02/2012/2012M020104524\\_S1Ver2.pdf](http://library.kaist.ac.kr/thesis02/2012/2012M020104524_S1Ver2.pdf).
- [2] GoK-EGS, e-Government Strategy 2004, Government press, 2004.

- [3] I. Syamsuddin, "Evaluation of e-government initiatives in developing countries: An ITPOSMO approach," International Research Journal of Applied and Basic Science, 2011.
- [4] GoK, Feasibility Study Report of Kenya Government Integrated Data Center, 2014.
- [5] R. Heeks, "Most e-Government for development projects fail: How can risks be reduced?" IDPM i-government working paper, 2003.
- [6] T. Altameen, M. Zairi and S. Alshawi, "Critical Success factors of e-government: A proposed model for e-government implementation," 2006.
- [7] J. Fortune and D. White, "Framing of projects critical success factors by a system mode," International Journal for project management, 2006.
- [8] R. Heeks, "e-government for development: Causes of e-government success and failure: Factor Model," IDPM, 2003a.
- [9] B. Spriano, "Success and failure of e-government projects in developing countries: The case of Zambia," 2013.
- [10] R. Heeks, "e-Government for development: Causes of e-government success and failure: Factor Model," IDPM, 2003.
- [11] A. Torki, Z. Mohamed and A. Sarmad, "Critical Success Factors of E-government: A Proposed Model for E-government Implementation," IEEEExplore, 2006.
- [12] N. Janja, "Change management as a critical success factor in e-government implementation," Journal of Business Systems Research, vol. vol 2, no. No 2, 2011.
- [13] A. M. Al-Naimat, S. M. Abdullah and M. A. Khairie, "Critical success factors for e-government implementation in Jordan," in 4th International conference on computing and informatics, Sarawak, 2013.
- [14] Mugenda and Mugenda, "research Methods," Nairobi Enon J.C (1998) Educational research, statistics and measurement, 2003.
- [15] A. Mohammed, S. Drew and R. AlGhamdi, "Analysis of Citizens' acceptance for E-government services: Applying the UTAUT Model," International conference theory and practice in modern computing and internet Application and research, 2012.
- [16] C. L. Kimberlin and A. G. Winterstein, "Validity and reliability of measurement instruments used in research," Am J Health-Syst Pharm, vol. 65, pp. 2276- 2284, 2008.
- [17] A. Field, "Reliability Analysis," in Research Methods II, 2006, p. 1.
- [18] R. Bentler and D. G. Bonett, "Significance Tests and goodness of Fit in the Analysis of Covariance Structures," pp. pp 37-51, 1988.
- [19] D. Gefen and D. Straub, "A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example," The communications of the Associations for Information Systems, vol. 16, no. 5, p. 91–109, 2005.
- [20] R. Bagozzi and Y. Yi, "n the Evaluation of Structural Equation Models," Journal of the Academy of Marketing Science, vol. 16, pp. 74-94, 1988.
- [21] D. Gichoya, "Factors affecting the success implementation of ICT projects in Government," The Electronic Journal of e-Government, vol. Volume 3, no. Issue 4, 2005.
- [22] GOK-EGS, e-government Strategy, 2004, Government press, 2004.
- [23] Heeks, Most e-government-for-development Projects fail, how can Risks be reduced? institute for Development Policy and Management, 2003.
- [24] R. Heeks, "Information Systems and developing countries: Failure, Success and Improvisations," Electronic Journal of e-government, 2002.
- [25] C. G. e. a. Hossan, "Success and failure factors for e-government projects implementation in Developing Countries: A study on perception of government official of Bangladesh," Electronic Journal of e-government, 2006.
- [26] N. Y. Lee, "E-government Application - The academy of ICT essentials for Government leaders," United Nations Asian and Pacific training entre for Information and Communication technology for development, 2009.
- [27] S. e. a. Shin, "Implementing e-government in developing countries: Its unique and common success factors," 2006.
- [28] A. Al-Naimat, M. Abdallah and F. Ahmad, "E-government implementation problems in developing countries," 2nd World Conference on Information Technology (WCIT-2011), vol. Volume 1 (2012), 2012.
- [29] R. Heeks, "Information Systems and developing countries: Failure, success and local improvisations," 2002.

# A Proposed Simulation Model of Automatic Machine For House Paint Selection Using Finite State Automata

Emigawaty  
Departement of Informatics  
STMIK AMIKOM  
Yogyakarta, Indonesia

---

**Abstract:** One important criterion considered by the public at the time of house selection or property is the aspect of house paint colors. Principally, the selection of house paint is not excessively complicated, but if the homeowner does not have enough knowledge about how to make a good combination with the paint color, the consequence is the color of choice is not in accordance with their expected. This study aims to provide a simulation of how to do a combination of paint color selection using automated machines. The method used in the design of this model is the Prototype method and blackbox testing to ensure the system is made valid and reliable. In the mean while, the approach automata models used are Finite State Automata (FSA) has the characteristics due to work by identifying and capturing the pattern in the compilation process of determining the color of wall paint the house. This paper also resulted in the identification of input symbols, FSA-NFA diagrams, rules production, NFA, equivalence transition table of FSA-DFA, and FSA-DFA transition diagram. At the same time, in the design of a prototype, also produced mockup simulation applications in order to generate automatic machine house paint selection. Thus, problems such as confusion and lack of knowledge of homeowners against a combination of paint colors can be overcome properly.

**Keywords:** automatic machine, prototype, finite state automata

---

## 1. INTRODUCTION

Along with the high expectations of today's human needs would be a challenge for developers of information technology to create models or equipment that can support daily activities. This is evidenced by the increasing number of variants of information technology innovation both based hardware and software which proved quite effective in helping humans for teaching and learning activities, work, sports, and other business aspects. One of the areas that affect the development of the information technology tools is the property sector. The rise of property business in almost all regions of the country is certainly a phenomenon as demand for property is also one of the primary needs of human. The developers of housing today competing to offer properties ranging from design models, thematic; to the color selection is also an important consideration to attract public attention.

The selection of paint color for some people may not be a priority compared to some other criteria. But not a few people have a high level of sensitivity to choose the appropriate color combinations or desired according to the taste or thematic expected by the owner of the house later. Determine the combination of colors of paint it takes precision and a good knowledge before deciding the color of choice. Because if we are incorrect in choosing and combining these colors is not likely to result from the combination does not conform to what is expected.

Automatic machine is one variant of the technology innovations created to change an activity or process that is the business of a conventional or to automatic with the goal of accelerating time activity process itself [1]. One of the automated machines that began to develop at this time is a compilation engine automatically selecting paint the walls of the house. The presence of the house paint compilation engine is expected to be one solution to simplify the electoral process and the determination of the color of paint with different variations. Finite State Automata (FSA) is an abstract machine in the form of mathematical models of systems with

discrete inputs and outputs that can recognize the simplest language also known as regular language and can be implemented in practice. FSA also serves as a model that is very useful to help recognizing and capturing a pattern in the data. This model can also receive inputs and produce outputs that have a finite number of states and can transform from one state to another based on input and transition functions in it [2]. FSA on a compilation machine house paint color selection is the right choice to model the process in this case. After simulating paint color selection is expected to house the homeowner will get a selection of color illustrations and recommendations in accordance with the needs and desires.

In this study, using a model of the application of the concept of FSA on the implementation of the automatic machine compilation paint of the house where the automatic machine has the characteristic assignment by identifying and capturing the pattern in the compilation process of determining the color of house paint. Therefore, problems such as confusion and lack of knowledge of homeowners against a combination of paint colors can be overcome with good. Moreover, this study also used some of the conditions that limit to do simulation and compiling election house paint colors using 3 size capacities gallon of paint (5kg, 10kg, and 15kg) with color inputs or raw are: yellow, blue, and red as primary colors, black and white as a neutral color, green (mixture of blue and yellow) as the secondary color, purple (mix of blue and red) as the secondary colors, orange (mix of red and yellow as secondary colors, while the composition used is 1 to 1.

## 2. LITERATURE REVIEW

### 2.1 Language Theory

In the process, a clump of computer science has two elements, namely: First, the model and the fundamental idea of the theory of computation. Secondly, the so-called engineering techniques for the system design of the computational model, both in terms of software and devices severity. Meanwhile, language and automata theory included in the first element



that makes this concept as a model and the fundamental idea of the theory of computation [3].

One theory being unique or singular when it may be good alternative solution of a phenomenon that exists in a science discipline. Computer science itself has a very broad scope ranging from concept, design theory, computer networks, programming, until the business processes within an information system. For that was created or constructed abstract models of the computer system or computing. Model theory of language has an important function, and generally how the computing system may be invoked on a software development and hardware [4].

In studying the theory of language, is identical with an understanding of formal language, primarily for the benefit of the design compiler and processing a script or text processor. By definition, a formal language is a set of sentences in a language that is raised by a grammar the same [3]. In its characteristics, that a formal language can be generated by two or more different grammar. Background that the theory of language can be called as well as the formal language for grammar created precedes the generation of each sentence. In contrast to human language, which is the reverse, while, grammar was created to formalize the words that live in the community.

Utdirartatmo [3] wrote that the automata is a system that consists of several finite state, where the state represents information regarding the input, and can also be translated as a reminder mesing. In this case, the input is meant automata machine is considered as the language that must be recognized by a machine. Then, the automata machines can make decisions that recognize or analyze whether the input is acceptable or not. In line with that rose by Linz [2], that for modeling hardware is a needed models automaton, which can also serve as a temporary storage medium and has the capability to transform an input to an output.

Furthermore, Juarna [5] provide some basic understanding of the language and automata theory as the basis of a common understanding of the definition and the main terminology as follows:

- Symbol is an abstract entity (such as the definition of a point in geometry). A letter or a figure is an example of the symbol.
- The string is a row of a limited (finite) symbols. For example, if a, b, and c are three symbols then abcb is a string that is constructed from the three symbols.
- If w is a string, the string length is expressed as  $|w|$  and defined for the count (number) symbols that make up the string. For example, if  $w = abcb$  then  $|w| = 4$ .
- The empty string is a string with zero fruit symbols. Empty string  $\epsilon$  is expressed by a symbol (or  $\wedge$ ) that  $|\epsilon| = 0$ . The empty string can be viewed as a symbol of a vacuum because both are composed of zero fruit symbols.
- Alphabet is himpunan up (finite set) symbols.

## 2.2 Grammar and Language

Grammar by Ullman [6] is a collection of the sets of variables, symbols terminal, the initial symbol, which is governed formally by the rules of production. Historically, teradapat a mathematician named Noam Chomsky in 1959 performed clustering or classification level language into four categories, which became known as Chomsky Hierarchy. The clustering can be seen in Table 1:

**Table 1. Classification Level of Language**

| Language   | Automatic Machine  | Production Rules Limitation   |
|--|--|---|
| Regular / Type 3   | Finite State Automata (FSA) covers Deterministic Finite Automata (DFA) & Non Deterministic Finite Automata (NFA) | $\alpha$ is a variabel symbol. Maximum $\beta$ has a variable symbol that when there is in the rightmost position |
| Context Free / Type 2                                      | Push Down Automata (PDA)   | $\alpha$ is a variable symbol   |
| Context Sensitive / Type 1                                 | Linier Bounded Automata  | $ \alpha  \leq  \beta $   |
| Unrestricted / Phase Structure / Natural Language / Type 0 | Turing Machine   | No Limitation   |

## 2.3 Finite State Automata (FSA)

Finite State Automata (FSA) or often also referred to as finite automata, in principle, not a physical machine but a mathematical model of a notation or phrase that can receive input and output is discrete. FSA is part of the machine automata regular languages, which have a finite number of states, and can move from one state to another. One shortcoming of these automata models is that the FSA does not have the storage allocation so that the ability to remember is very limited [3].

Furthermore, as proposed by Ullman [7] that the FSA is a tool that can be used to design a system, whereby the mechanism can be applied or implemented on lexical analysis, text-editor, the network communication protocol, and checks parity, elevator system, traffic light system, or also the automatic machine. FSA formally expressed by 5 tuple or  $M = (Q, \Sigma, \delta, S, F)$ , where:

- Q = set of state / status
- $\Sigma$  = set of symbols input / input / alphabet
- $\delta$  = transition function
- S = the initial state / initial position (initial state)
- F = set the final state

Finite State Automata is a machine that can recognize regular language classes and has the following characteristics:

- Ribbons inputs contains a series of symbols (string) from the set of symbols / alphabet,
- Every time after reading one character, posisir read head will be in the next symbol,
- At any time, the FSA is at a certain status,
- The number of valid status for the FSA is finite.

Based on the function and definition of Finite State Automata (FSA) and also in terms of its ability to change its state then this model can dikasifikasikan into 2 groups: Deterministic Finite Automata and Non-Deterministic Finite Automata.

## 2.4 Deterministic Finite Automata (DFA)

At Deterministic Finite Automata (DFA), from a state there is exactly one next state for each symbol of the input received, cannot be empty and more than one state. DFA formally expressed by 5 tuple or  $M = (Q, \Sigma, \delta, S, F)$ , where Ullman [7]:

- Q = set of state / status
- $\Sigma$  = set of symbols
- $\delta$  = transition function
- S = the initial state ( $S \in Q$ )
- F = set the final state (can be more than 1)



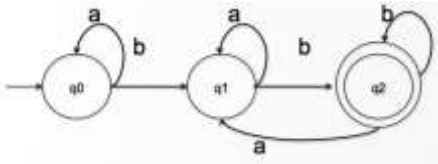


Figure 1. Deterministics Finite Automata (DFA) Machine

DFA configuration in Figure 1 is expressed formally as follows:

- Q = {q0, q1, q2}
- $\Sigma$  = {a, b}
- S = q0
- F = q2

While the transition of existing functions in Figure 1 are as follows:

- d(q0,a) = q0
- d(q0,b) = q1
- d(q1,a) = q1
- d(q1,b) = q2
- d(q2,a) = q1
- d(q2,b) = q2

## 2.5 Non-Deterministic Finite Automata (NFA)

In the non-deterministic finite automata of a state there can be 0,1, or more bows out (transition) labeled with the same input symbol. FSA formally expressed by 5 tuple or  $M = (Q, \Sigma, \delta, S, F)$ , where:

- Q = set of state / status
- $\Sigma$  = set of symbols
- $\delta$  = transition function
- S = the initial state ( $S \in Q$ )
- F = set the final state (can kebih of 1)

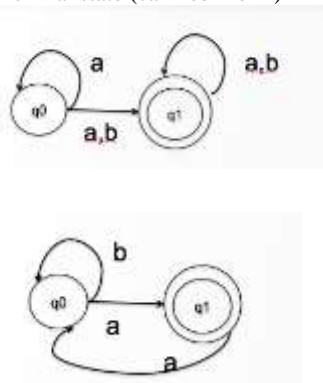


Figure 2. Non-Deterministics Finite Automata (DFA) Machine

## 2.6 Related Works

1. Melly [8] with the title of paper "Application of Concept Finite State Automata at Coffee Beverage Maker Machine Auto". In this study focused on an application of the concept of FSA in an automatic coffee machine beverage maker. The method used in this study is the implementation of formal methods that can be used to bridge (the formal specification) the manufacture,

development and verification of hardware and software, which can be used from the initial design to the test results. The study produced a simulation applications beverage automatic coffee maker is an application simulating a machine that can make the process of making coffee drinks and variations automatically where the completion of the process used concept of FSA. Disadvantages of this study are not yet modeled Equivalence of NFA to DFA thus cannot be recommended to be developed or translated into the programming language.

2. Suprihatin [9], titled paper "Finite State Automata for Parsing Beheading Syllabary In Indonesian". In this study how decapitated syllables in Indonesian is good and right, while the FSA itself functions as an abstract machine, which is used to help design, the parsing (decapitation) syllable. FSA in its implementation is used to describe the structure of vowels, consonants, or the structure of words in Indonesian accordance with Spelling Enhanced (EYD). The downside of this research has not been able to show in detail the transition diagram for each of the components or variables measured.
3. Dihya [10], with paper title "Implementation of the Transitional Graf Defining Formal Languages". This study on how to model a formal language with transition graph and proof when modeled with other modeling. This is done to determine how the machine answers the problem decisions based on the input string to provide the output of 'yes' or 'no'. So that it can be seen how the issues will be seen run private computing simple graphical representation of the transition. The downside of this research is only at the stage of analysis and produces a transition table, not to the implementation of an object or case studies, so the validity of this study cannot be measured.

## 3. METHODOLOGY

### 3.1 Prototype Model

Methodology is the unity of methods or rules used by the work of a science. While the method is a systematic method or technique to grind. The general objective of this development is to provide ease in conveying information, reduce costs and save time, improve control, drive growth, improve the productivity and profitability of the organization.

System development can mean the preparation of a new system to replace the old system as a whole or improve existing systems. System development method used is the method Prototype. Prototype is a method used in a systems approach to make something quick and gradual program that can immediately be evaluated by the user. The stages contained in this prototype method is as follows [11]:

1. Identification of User Needs, at this stage developers and users meet. User describes the system requirements.
2. Create Prototype; developers began to create a prototype of the system.
3. Test Prototype, after prototype form testing prototype user and give criticism or suggestions.
4. Improving Prototype, In This stage developers make modifications in accordance with the input from the user.
5. Develop Prototype, after evaluation and perfect the system in accordance with the wishes of the user. Then the developer completed the final system in accordance with the insert of the wearer.

### 3.2 FSA Formal Specification

Formal specification by definition is a specification or statement, which is represented in the form of a formal language with the intention to describe what, should be done or decided by the software [12]. In the implementation, this study will use a formal specification in order to make a model of an object or system behavior using mathematical approaches such as sets and sequences.

After having a specification, either in the form of formal or informal, the implementation can be made. Implementation of the model can be made of conventional or automatic. Making the manual implementation is only suitable for use in the design of the level of complexity is not too high. For complex systems, easy error occurred. The process of making the implementation is automatically called automatic synthesis. To guarantee the truth of the synthesis process, the synthesis process must be substantiated. If a tool for automatic synthesis and synthesis processes that are used can be proved right then the results can be said to be "correct by construction".

Usually this is done to a system that departs from a system specified formally, then performed "refining" in order to obtain results in more detail, and done repeatedly until it becomes implementations in the level of abstraction desired (eg up to level layout transistor). Each step in the refining process is guarded by a mechanism (for example, by a theorem prover) so it cannot perform the process of refining the wrong [12].

## 4. RESULTS AND DISCUSSION

### 4.1 Input Symbol Identification

In the design stage and construction of the simulation modeling of automated machinery house paint color selection using the FSA, the necessary step towards the identification of input symbols in advance. This is necessary in order to provide the initial terms of the initiation of the formal specification by way of input variables, process variables, and variable selection. In practice, this study will use a formal specification that is used to create a model of an object or system behavior using arithmetic approach such as sets and sequences.

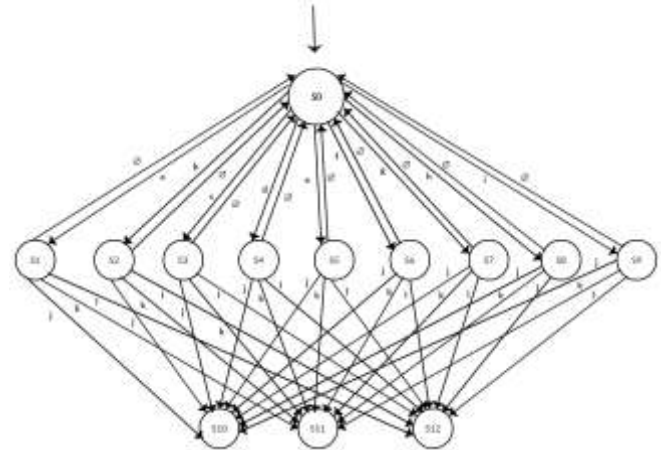
The following is a formal specification through initiation illustration of input and output variables:

**Table 2. Formal Specification of Input Identification**

| Process Initiate      | Capacity Selection | Compile Process |
|-----------------------|--------------------|-----------------|
| 0 Back to start state | j Gallon 5 kg      | m Red           |
| a Choose green        | k Gallon 10 kg     | n Add Red       |
| b Choose light green  | l Gallon 15 kg     | o Yellow        |
| c Choose dark green   |                    | p Add Yellow    |
| d Choose purple       |                    | q Blue          |
| e Choose light purple |                    | r Add Blue      |
| f Choose dark purple  |                    | s Black         |
| g Choose orange       |                    | t Add Black     |
| h Choose light orange |                    | u White         |
| i Choose orange tua   |                    | v Add White     |
|                       |                    | w Compile       |

In Table 2 describes the formal specification in the initiation input variables in the form of a variable to i, and the process variables, namely m up to w. As for the election process capacity is represented by the variable j, k, and l. In the implementation, application built will have an interface to allow the user or the user selects multiple colors and desired gallon capacity according to their needs.

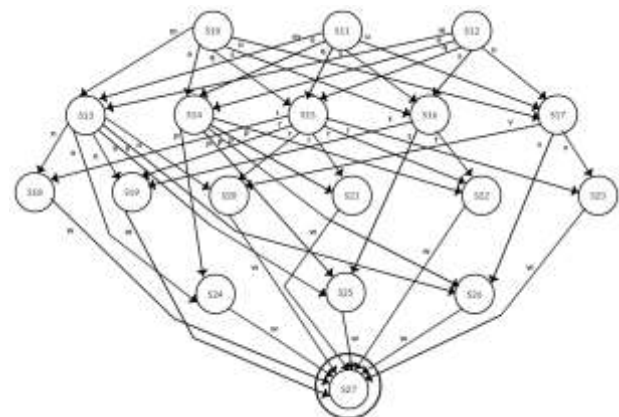
### 4.2 FSA – Non DFA



**Figure 3. Diagram of Finite State Automata - NFA**

In Figure 3 it can be seen that the Finite State Automata start work on State S0, in this diagram, there are 9 colors that can be, namely: green [a] to State S1, light green color [b] to State S2, dark green color [c] to State S3, purple [d] to the State S4, light purple [e] to the State S5, and petunia [f] to State S6. Then, from State to State S1 to S9 choice for gallon capacity, ie paint 5kg [j] to State S10, paint 10kg [k] to State S11, and paint 15kg [l] to State S12. Mathematical model on FSA-NFA diagram is then translated into the design of the interface by providing drop down option or combo box for each option.

Transition specification diagram illustrates the process that is automated election house paint using the concept or theory of FSA. In principle, the machine is designed to mengikukui pattern of the flow of paint color selection process is based on several primary colors or base is then processed according to color choices finally as an output. Is because each state has had towards both the source state and state clear objectives, then it is highly unlikely to happen mistakes or errors in processing or even have an end result that is not in accordance with a predetermined scheme. In the mean time, in order to be able to read and interpret symbols input the formal specification designed, the FSA granted terminology in the form of a dictionary of symbols is fixed and redundancy with other processes.



**Figure 4. Diagram of Finite State Automata - NFA (continued)**

Figure 4 is an advanced process flow previously, where the State S1 to S9 State 5kg choice (j) to State S10, 10kg (k) to State S11 and 15kg (l) to S12. Furthermore, from State S10, S11, and S12, if the choice of red (m) to State S13, yellow (o) to State S14, blue (q) to State S15, black (s) to State S16, and

white (u) to State S17. After that, from State S13 add red (n) to State S18, S19, S20, S24, S25, and S26. Of State S14 add yellow (p) to State S21, S22, S23, S24, S25, and S26. Of State S15 add blue (r) to S18, S19, S20, S21, S22, and S23. Of State S16 add black color (s) to State S19, S22, and S25. Of State S17 add white (v) to State S20, S23, and S26. Furthermore, from State S18 to State S26 stirring (w) to State end S27. From State to State S1 to S27 option  $\emptyset$  (cancellation) back to S0.

### 4.3 Rule Production of NFA

In the domain of language and automata theory, grammar can also be referred to as Grammar, which is a collection of the set of variables, symbols terminal (start and an end state), the initial symbols are constrained by the rules of production. Furthermore, the production rules in a State Non-Deterministic Finite Automata (NFA) have produced some string variables.

NFA Production Rules:

S0 = S  
 S1 = A S2 = B S3 = C S4 = D  
 S5 = E S6 = F S7 = G S8 = H  
 S9 = I S10 = J S11 = K S12 = L  
 S13 = M S14 = N S15 = O S16 = P  
 S17 = Q S18 = R S19 = S S20 = T  
 S21 = U S22 = V S23 = W S24 = X  
 S25 = Y S26 = Z  
 S27 = F

Thus,

G = {VT, VN, S, P}  
 VT = {a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w, $\emptyset$ }  
 VN = {A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z}  
 S = S

P = { S  $\rightarrow$  aA | bB | cC | dD | eE | fF | gG | hH | iI |  $\emptyset$ ,  
 A  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 B  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 C  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 D  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 E  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 F  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 G  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 H  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 I  $\rightarrow$  jJ | kK | lL |  $\emptyset$   
 J  $\rightarrow$  mM | oN | qO | sP | uQ |  $\emptyset$   
 K  $\rightarrow$  mM | oN | qO | sP | uQ |  $\emptyset$   
 L  $\rightarrow$  mM | oN | qO | sP | uQ |  $\emptyset$   
 M  $\rightarrow$  nR | nS | nT | nX | nY | nZ |  $\emptyset$   
 N  $\rightarrow$  pU | pV | pW | pX | pY | pZ |  $\emptyset$   
 O  $\rightarrow$  rR | rS | rT | rU | rV | rW |  $\emptyset$   
 P  $\rightarrow$  tS | tV | tY |  $\emptyset$   
 Q  $\rightarrow$  nT | vW | vZ |  $\emptyset$   
 R  $\rightarrow$  w | S  
 S  $\rightarrow$  w | S  
 T  $\rightarrow$  w | S  
 U  $\rightarrow$  w | S  
 V  $\rightarrow$  w | S

W  $\rightarrow$  w | S  
 X  $\rightarrow$  w | S  
 Y  $\rightarrow$  w | S  
 Z  $\rightarrow$  w | S

### 4.4 Table of Transition Equivalences

Basic concept of Finite State Automata (FSA) provide insight into fundamental that when a system is in a state of a state that moves between state-state that no principle can be produced depends on the input to the system. It is confirmed that the input (input) is true and valid is crucial for conditioned value without error. Implementation of this concept can be seen in modeling compilation or translation of high-level programming language (high level language) into machine language equivalent.

From Non-Deterministic Finite Automata (NFA) machine, which had previously modeled, then the machine can be made Deterministic Finite Automata (DFA) equivalent. Equivalent is meant accords or able to accept the same language. In the case of this study, the transition equivalence can be seen in Table 4.2. The table contains a description of all State with input option (input) respectively to the state-state is written.

Table 3. Equivalency Transition from NFA to DFA

### 4.5 FSA- DFA Transition Diagram

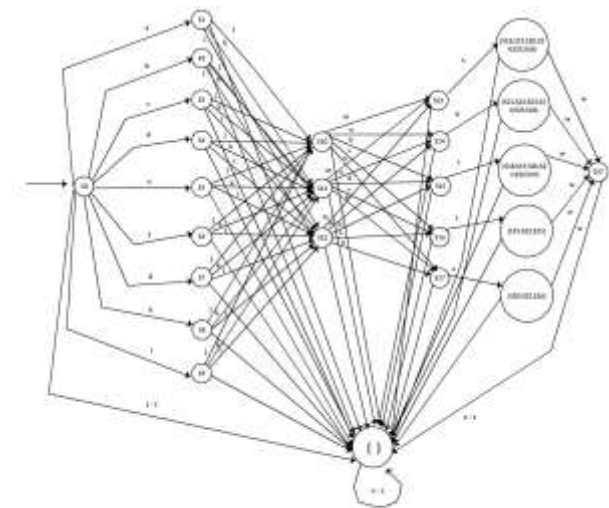


Figure 5. Finite State Automata – NFA Transition Diagram

In Figure 5 can be explained that the FSA began work on the State S0, from S0 option A to S1, b to S2, c to S3, d to S4, e to S5, f to S6, g to S7, h to S8, i to S9, j - z State to an empty ({}). Then, from the State S1 to S9 to S10 option j, k and l to S11 to S12, to a - i and m - z State to an empty ({}).

Furthermore, from State S10, S11, and S12 option m to S13, o to S14, q to S15, s to S16 and u to S17, for a - l, n, p, r, t, v - z to State empty ,

Afterwards, from State to State S13 option n new ({SA18, S19, S20, S24, S25, S26}). For a - v, x - z to State empty. From State S14 selection to p ({S21, S22, S23, S24, S25, S26}). For a - o, q - z to State empty. From State S15 selection to r ({S19, S20, S21, S22, S23}). For a - q, s - z to State empty. State S16 selection of t to ({S19, S22, S25}). For a - s, v - z to State empty. From State S17 to choice v ({S20, S23, S26}). For a - v, w - z to State empty. Of the newly formed State 5 w option to S27. For a - v, x - z to State empty. And for all the existing state S0 option other than 0 to S0.

#### 4.6 Application Prototype



Figure 6. Prototype of Paint Color Selection: Green

Figure 6 illustrates the main interface where the user can immediately make paint color selection, and also the capacity of gallons of paint in accordance with the wishes or needs. In this example, the user make the selection with the green color gallon capacity is 5kg, and a button to execute the process of mixing paint.

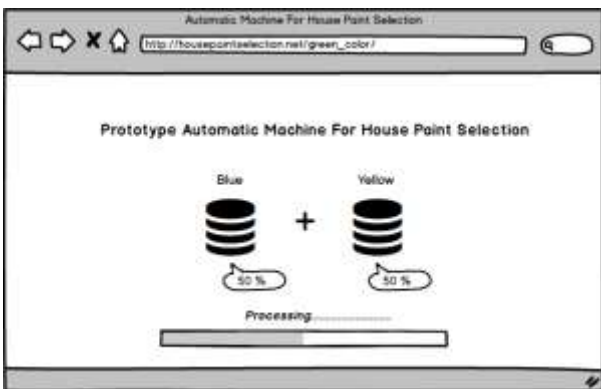


Figure 7. Prototype of Automated Paint Mixing Process: Green

Meanwhile, in Figure 7 can be seen that the application displays the information about the process through simulation where the system will demonstrate the process of stirring or mixing paint, the color blue (the portion of the composition 50%) plus Yellow (the portion of the composition of 50%) to produce a green color.

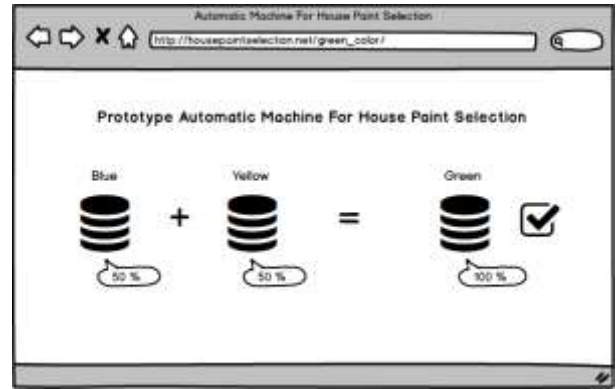


Figure 8. Prototype of Automated Paint Mixing Final Result: Green

Figure 8 is a display end result of the process of stirring or mixing colors to produce a green color. In this application also displays the composition of each color to help users determine the proportion of each paint.

#### 4.7 Prototype Testing

As an important part of assessing the success of a design prototype, so in this study is also testing a system to produce valid and reliable. There are several methods in testing of the prototype as black box testing, white box testing, or usability testing. But in this study using black box testing, the testing model by observing the process inputs (input) and output (output) in the system software to ignore what is going on in the system.

Table 4. Blackbox Prototype Testing

| No | Indicators         | Scenario   | Expected Output   | Obtained Output  |
|----|--------------------|--|---|--|
| 1  | Form Design        | User run applications from the beginning to the end (to get the paint color desirably) | <ul style="list-style-type: none"> <li>Colors selection form</li> <li>Paint mixing process form</li> <li>Result paint selection form</li> </ul> | <ul style="list-style-type: none"> <li>Acceptable</li> <li>Acceptable</li> <li>Acceptable</li> </ul> |
| 2  | Navigation         | User run applications from the beginning to the end (to get the paint color desirably) | Button process and navigation system for each forms   | Acceptable   |
| 3  | Respon             | User run applications from the beginning to the end (to get the paint color desirably) | Duration of response of applications ranging from input to final results  | Acceptable   |
| 4  | Transition Diagram | User run applications from the beginning to the end (to get the paint color desirably) | Concurrence between output with grammar and FSA Transition Diagram - NDFAs  | Acceptable   |



## 5. CONCLUSION

Based on the problems and objectives of this study, it can be delivered several conclusions as follows:

1. This research resulted in Diagram Finite State Automata - Non Deterministic Finite State Automata, Production Rules, Transition Equivalence, and Diagrams Finite State Automata - Deterministic Finite State Automata, also designed a prototype of an automatic machine simulation application selection of paint house as one of the models implementation.
2. Based on the results of the testing of a prototype application, it can be seen that the concept of Finite State Automata (FSA) used in designing an automatic machine simulation application selection with house paint to generate a grammar (grammar) to produce nine color options can house paint proved successful according to the research objectives.

## REFERENCES

- [1] Melly, R. I. (2012). Penerapan Konsep Finite State Automata (FSA) Pada Mesin Pembuat Kopi Otomatis. *Jurnal komputasi*, 1, 95-102.
- [2] Linz, P. (2001). *An Introduction to Formal Language and Automata*. John and Bartlett Publisher.
- [3] Utdirartatmo, F. (2005). *Teori Bahasa dan Otomata*. Yogyakarta: Graha Ilmu.
- [4] Kelley, D. (1999). *Otomata dan Bahasa-bahasa Formal*. Prenhanindo.
- [5] Juarna, A. (2008). *Teori Bahasa dan Automata*.
- [6] Ullman, J. E. (1979). *Introduction to Automata Theory, Languages & Computation*. Addison Wessley Inc.
- [7] Ulman, A. A. (1972). *The Theory of Parsing, Translation and Compilling* (Vol. I). Prentice Hall.
- [8] Melly, R. I. (2012). Penerapan Konsep Finite State Automata (FSA) Pada Mesin Pembuat Kopi Otomatis. *Jurnal komputasi*, 1, 95-102.
- [9] Suprihatin. (2013). *Finite State Automata Untuk Parsing Pemenggalan Suku Kata Dalam Bahasa Indonesia*. Yogyakarta: Universitas Negeri Yogyakarta.
- [10] Dihya, A. (2010). *Penerapan Graf Transisi dalam Mendefinisikan Bahasa Formal*. Bandung.
- [11] McLeod, R. &. (2004). *Sistem Informasi Manajemen Edisi Kedelapan*. Jakarta: PT Indeks.
- [12] Raharjo, B. (1998). *Penggunaan Formal Methods dalam Desain Perangkat Keras*. Pusat Penelitian Antar Universitas bidang Mikroelektronika (PPAU ME) Institut Teknologi Bandung.



# Redundant Actor Based Multi-Hole Healing System for Mobile Sensor Networks

S. Himabindu  
VBIT  
Hyderabad

B. Brahma Reddy  
VBIT  
Hyderabad

## Abstract

*In recent years, the Mobile Wireless Sensor Network is the emerging solution for monitoring of a specified region of interest. Several anomalies can occur in WSNs that impair their desired functionalities resulting in the formation of different kinds of holes, namely: coverage holes, routing holes. Our ultimate aim is to cover total area without coverage hole in wireless sensor networks. We propose a comprehensive solution, called holes detection and healing. We divided our proposed work into two phases. The first phase consists of three sub-tasks; Hole-identification, Hole-discovery and border detection. The second phase treats the Hole-healing with novel concept, hole healing area. It consists of two sub-tasks; Hole healing area determination and node relocation.*

**Key words:** WSN, coverage holes, routing holes, hole healing.

## 1. INTRODUCTION

RECENT years have witnessed a growing interest in the applications of wireless sensor-actor networks (WSANs). Of particular interest are applications in remote and harsh areas in which human intervention is risky or impractical. Examples include space exploration, battle field surveillance, search-and-research, and coastal and border protection. A WSAN consists of a set of miniaturized low-cost sensors that are spread in an area of interest to measure ambient conditions in the vicinity.

Given the collaborative actors' operation, a strongly connected inter-actor network topology would be required at all times. Actors usually coordinate their motion so that they stay reachable to each other. However, a failure of an actor may cause the network to partition into disjoint blocks and would thus violate such a connectivity requirement.

## 2. RELATED WORK

The [1] wireless communications systems in the mobile computers support a broadcast mechanism, much more flexible and useful ways of sharing information can be imagined. Our proposed routing method allows a collection of mobile computers, which may not be close to any base station and can exchange data along changing and arbitrary paths of interconnection, to afford all computers among their number a (possibly multi-hop) path along

which data can be exchanged. In addition, our solution must remain compatible with operation in cases where a base station is available. By the methods outlined in this paper not only will routing be seen to solve the problems associated with ad-hoc networks, but in addition authors will describe ways to perform such routing functions at Layer 2, which traditionally has not been utilized as a protocol level for routing. Packets are transmitted between the stations of the network by using routing tables which are stored at each station of the network. Each routing table, at each of the stations, lists all available destinations, and the number of hops to each. Each route table entry is tagged with a sequence number which is originated by the destination station. To maintain the consistency of routing tables in a dynamically varying topology, each station periodically transmits updates, and transmits updates immediately when significant new information is available, since authors did not assume that the mobile hosts are maintaining any sort of time synchronization, authors also make no assumption about the phase relationship of the update periods between the mobile hosts. These packets indicate which stations are accessible from each station and the number of hops necessary to reach these accessible stations, as is often done in distance-vector routing algorithms.

In [2] report, authors focus on the specific problem of altering the positions of robots in order to achieve a desirable ad hoc network topology starting from an arbitrary initial spatial configuration. Every robot node includes its location information (GPS coordinates or indoor relative location information) whenever it floods an LSU (link state update) to the rest of the network. Additionally neighbour information of a node is also extracted from an LSU in order to construct a view of the current network topology.

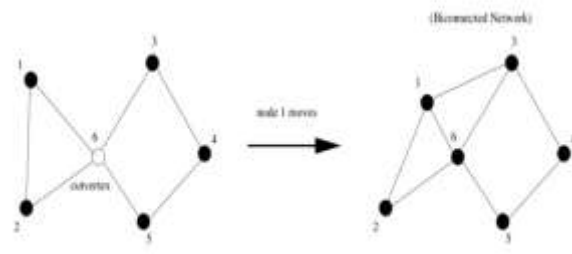


Figure 1: Achieving Biconnectivity by Node Movement

In [2] article authors show that iterative block movement algorithm significantly. Authors have considered a bi-connected network where nodes still can exchange messages with each other to coordinate the recovery process even after failure. And finding an exact polynomial time optimization algorithm for the 2D case is extremely hard. [3] Paper considers the connectivity restoration problem subject to path length constraints. Basically, in some applications, such as combat robotic networks and search-and-rescue operation, timely coordination among the actors is required, and extending the shortest path between two actors as a side effect of the recovery process would not be acceptable.

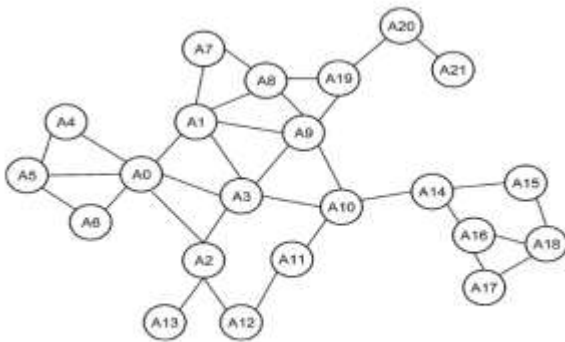


Fig. 1. Example one-connected inter-actor network. Nodes  $A_0$ ,  $A_{10}$ ,  $A_{14}$ , and  $A_{19}$  are cut vertices whose failure leaves the network partitioned into two or multiple disjoint blocks.

In [4] paper, authors designed and evaluated distributed self-deployment protocols for mobile sensors. After discovering a coverage hole, the proposed protocols calculate the target positions of the sensors where they should move. Author problem statement is: given the target area, how to maximize the sensor coverage with less time, movement distance and message complexity. In [4] paper authors addressed the problem of placing sensors in target field to maximize the sensing coverage. Although the centralized approach may minimize the sensor movement, central server architecture may not be feasible in some applications.

## 2.1 Existing system & Disadvantage

In existing research works, the researchers proposed the technique with movement assisted sensor node deployment and then Hole-detection works by a node decides whether it is on the boundary of a hole by comparing its degree with the average degree of its 2-hop neighbors. Not all boundary nodes can be identified correctly by this algorithm.

## 3. PROPOSED SYSTEM

Our proposed hole and border detection algorithm is distributed and lightweight, and thus more suited to the energy constrained

WSNs. we propose collaborative mechanisms to detect and heal holes. Our hole-detection mechanism deals with holes of various forms and sizes. We try to alert a limited number of nodes surrounding the hole, only those nodes have the task of moving and repairing the hole. In this section, we are going to discuss about our enhancement work. Our base method works like reactive mode, if the node is failed then only Hole healing will be start. By our base work we can cover the holes, but reactive mode will be cause to high level topology changes, and the more number of nodes has to move from own position. Due continuous node failure, the network may not be healed after certain healing process. To avoid this type of problem, we propose the extra temporary SensRob's. in our Enhancement we propose the failure detection based on the energy loss. The node will fail when it loosed remaining energy less than critical level. So in our enhancement, we propose the technique to detect failure of node due to the energy loss.

Each node in the network monitors the own energy loss, if own energy is getting reduce near to critical level then the node will inform to the base station about energy loss. Once base station received error message from the sensor the it will check availability of extra actor. The base node will share the position information of critical node and id details to the extra actor, and then the extra actor has to move to the critical node position.

After reaching the place of critical node, the extra actor has to inform to the critical node to inter change the id in neighbor's neighbor table. The critical node will send the inter change message to neighbor with extra actor id. The neighbors of critical node need to change their neighbor table info such as extra actor id will be placed as the neighbor sensor and neighbor sensor (critical node will be deleted). After id inter changing process, the critical node will be treated as extra actor and extra actor will be treated as normal sensor device, then the extra actor node (inter changed critical node) has to move to base station and its need to repair the energy level. This reclaimed extra actor can be placed in any other critical node in future.

## 3.1 Algorithm

### Hole detection & Healing :

- 1) **Initialize the Htimer and Neigh\_timer**
- 2) **If Timer expire**
  - **Generate the hello message**
    - $Pkt \leftarrow nd_{id} \& Pos(x, y)$
  - Broadcast Hello message
  - $Schedule(T_{now} + T_{Rand_{const}})$
- 3) **If Pkt Recv in node in  $n_i$** 
  - If  $Pkt.type = Hello$ 
    - Set  $Insert = true$
    - **Foreach  $M \in Tbl_N$**

- If  $Pkt.src = M.Id$ 
    - $\wedge M.T_{ex} = T_{now} + T_{const}$
    - $\wedge M.Pos(Pkt.x, Pkt.y)$
    - $\wedge Set\ Insert = false$
    - $\wedge Search_{stop}$
  - If  $Insert = True$ 
    - Allocate new Memory  $M \rightarrow Tbl_N$ 
      - $\wedge M.Id = Pkt.src$
      - $\wedge M.Pos(Pkt.x, Pkt.y)$
      - $\wedge M.T_{ex} = T_{now} + T_{const}$
  - Else-If  $Pkt.type = Hole_{disc}$ 
    - If  $Pkt.fid \in Nb_{list} \in Nb_{table} \ \& \ n_i \notin Pkt.path \ \& \ Pkt.src \neq n_i$ 
      - Set  $dst = Dist(Pkt.SPos_{x,y} \ \& \ n_{i(x,y)})$
      - If  $dst > Pkt.dst_{mx}$ 
        - $\wedge Pkt.dst_{mx} = dst$
        - $\wedge Pkt.Cpos \leftarrow (x_{n_i}, y_{n_i})$
      - $n_i \cup pkt.path$
      - $Reset(Timer_{Hole_{disc}})$
      - Set  $Pkt.hop = 0$
      - $Rebroadcast(Pkt)$
    - Else-If  $Pkt.fid \neq Id_{fail} \ \& \ n_i \notin Pkt.path \ \& \ Pkt.src \neq n_i \ \& \ Pkt.hop < 1$ 
      - $Pkt.hop ++$
      - $Rebroadcast(Pkt)$
    - Else-If  $Pkt.src = n_i$ 
      - If  $MDist_{fail} < Pkt.dst_{mx}$ 
        - $\wedge Set\ N.type = HM$
        - $\wedge MDist_{fail} = Pkt.dst_{mx}$
        - $\wedge x_c = (x_{n_i} + Pkt.Cpos_x)/2$
        - $\wedge y_c = (y_{n_i} + Pkt.Cpos_y)/2$
      - $Free(pkt)$
      - $Timer_{Hole_{Heal}}.Schedule(T_{now} + T_{Rand_{const}})$
    - Else
      - $Free(pkt)$
  - Else-if  $Pkt.type = Hole_{Heal}$ 
    - If  $Near_{hole} \ \& \ n_i \notin Pkt.path \ \& \ Pkt.src \neq n_i \ \& \ (\exists Lnk_{n_i, Pkt.Path_{i-1}} \ || \ \exists Lnk_{(H_c, Pkt.Path_{i-1})})$ 
      - If  $(Pkt.\frac{dst}{2}) > Dist(Cpos_{x,y} \ \& \ (x_{n_i}, y_{n_i}))$ 
        - $\wedge Set\ Future_{pos} \leftarrow \frac{Cpos_{x,y} + (x_{n_i}, y_{n_i})}{2}$
        - $\wedge Mov(Future_{pos})$
      - $Rebroadcast(Pkt)$
    - Else if
      - If  $!Near_{hole} \ \& \ n_i \notin Pkt.path \ \& \ Pkt.src \neq n_i \ \& \ \exists Lnk_{(n_i, Pkt.Path_{i-1}, Ftr_{x,y})}$ 
        - $\wedge Move_{(100 \rightarrow Pkt.Path_{i-1})}$
    - Else
      - $Free(pkt)$
- 4) **If Node  $n_i$ 's  $Timer_{Neig}$  expired**
- **For each  $M \in Tbl_N$** 
    - If  $M.T_{ex} < T_{now}$ 
      - Set  $Id_{fail} = M.Id$
      - $Delete(M)$
    - $Timer_{Hole_{disc}}.Schedule(T_{now} + T_{Rand})$
  - $Schedule(T_{now} + T_{Rand})$
- 5) **Node  $n_i$ 's If  $Timer_{Hole_{disc}}$  expired**
- Generate  $pkt$ 
    - $MDist_{fail} = 0$
    - $n_i \rightarrow pkt.src$
    - $pkt.type = Hole_{disc}$
    - $(x_{n_i}, y_{n_i}) \rightarrow pkt.Spos$
    - $n_i \rightarrow pkt.path$
    - $Init(pkt.hop \ \& \ pkt.Cpos)$
  - $broadcast(pkt)$
- 6) **If  $Timer_{Hole_{Heal}}$  expired in Node  $n_i$**
- Generate  $Pkt$ 
    - $Pkt.src = n_i$
    - $(x_c, y_c) \rightarrow Pkt.Cpos$
    - $MDist_{fail} \rightarrow Pkt.dst$
    - $Pkt.type = Hole_{Heal}$
  - $broadcast(Pkt)$

### 3.2 Pre-Failure Controller

Let,  $E_c$  for Current energy level,  $E_{Th}$  for threshold energy level,  $L_{critic}$  for critical node list,  $L_{Exact}$  for Available Extra actor list,  $Id_{Ex}$  for Extra actor Id,  $Pos$  for postion,

- 1) If  $E_c < E_{Th}$ 
  - a. Generate  $Pkt.critical$
  - b.  $Pkt.Nd = N_{id}$
  - c. Broadcast  $Pkt$
- 2) If  $Pkt$  recv in  $N$ 
  - a. If  $pkt$  is Duplicate
    - i. Free  $Pkt$
    - ii. Return
  - b. If  $Pkt.critical$ 
    - i. If  $N \neq BS$ 
      1. Rebroadcast  $Pkt$
    - ii. If  $N = BS$ 
      1.  $Pkt.Nd \cup L_{critic}$
      2. If  $L_{Exact} \neq Null$ 
        - a.  $Id_{Ex} = L_{Exact}(1)$
        - b.  $Rearrange(L_{Exact})$
        - c.  $Move(Id_{Ex} \rightarrow L_{critic}(1).Pos)$
        - d.  $Rearrange(L_{critic})$
  - c. If  $Pkt.Exact_{arrive}$ 
    - i. If  $N \neq Nd_{critical}$ 
      1. Ignore( $Pkt$ )
      2. return

- ii.  $Switch_{neigh}(N \rightarrow Id_{Ex})$
- iii.  $Move(N \rightarrow BS.pos)$

## 4. REQUIREMENTS

**Hardware:** Single PC with 20 Gb Hard disc space 1Gb RAM

**Software:** Linux OS (Ubuntu 10.04), NS2.34

**Languages:** TCL (Front end type project only), C++ (**Optional**)

- **Result**

We have tested our output with ns2 simulator and we got a two results, one is NAM, Xgraph.

Our enhancement method provides best results such as no node failure and less movement.

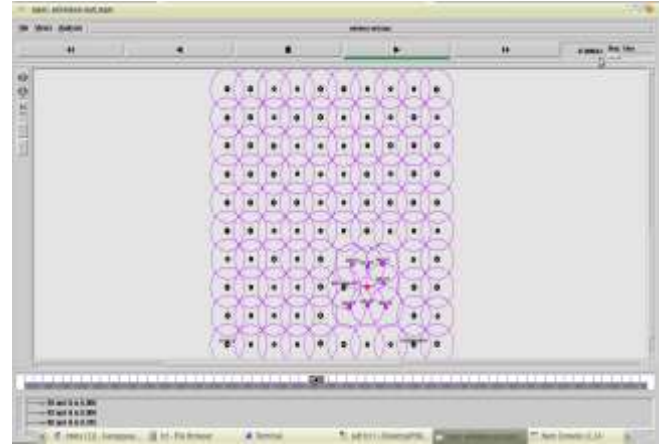


Fig. A3 Hole healing

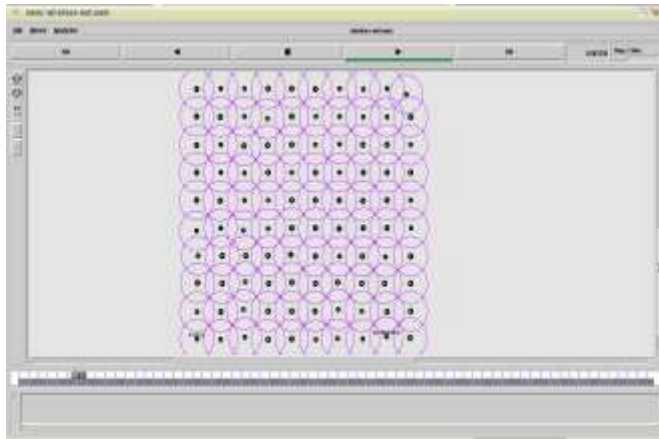


Fig. A1 Network placement and sensing area

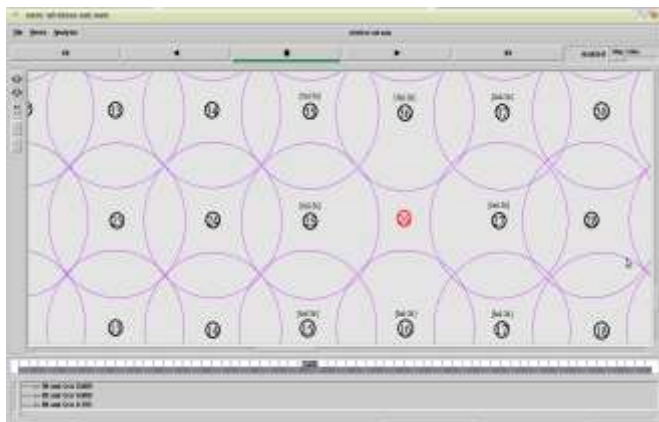


Fig. A2 Hole detection

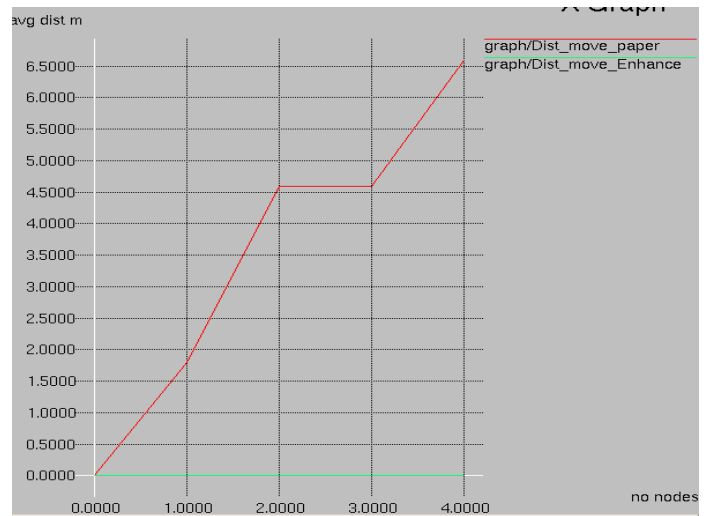


Fig.A4 Graph for avg moving distance

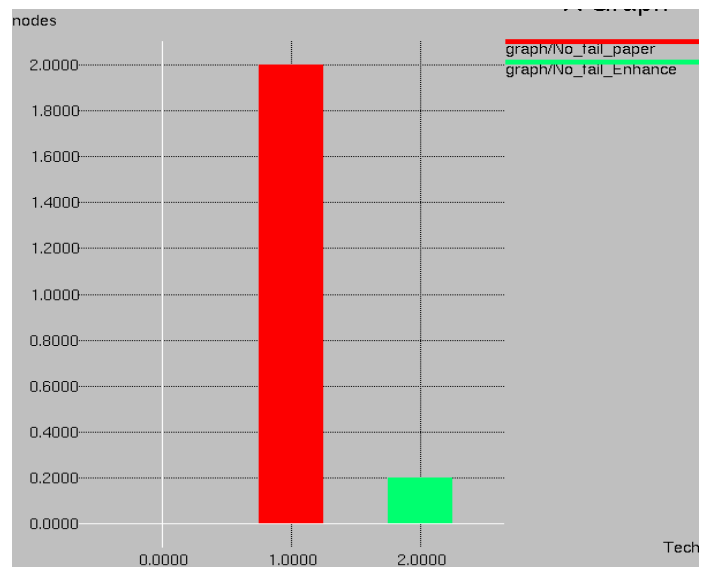


Fig.A5 Node failure

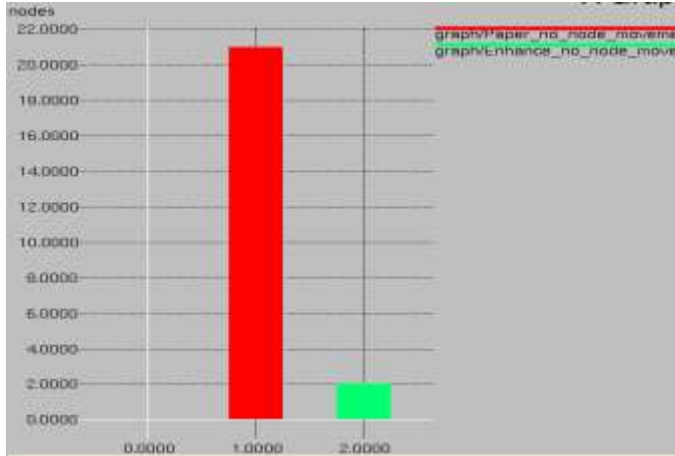


Fig.A6 Node movement

## 5. CONCLUSION

The Mobile Wireless Sensor Network is the emerging solution for monitoring of a specified region of interest. Several anomalies can occur in WSNs that impair their desired functionalities resulting in the formation of different kinds of holes, namely: coverage holes, routing holes. Our ultimate aim is to cover total area without coverage hole in wireless sensor networks. We propose a comprehensive solution, called holes detection and healing. We divided our proposed work into two phases. The first phase consists of three sub- tasks; Hole-identification, Hole-discovery and border detection. The second phase treats the Hole-healing with novel concept, hole healing area. It consists of two sub-tasks; Hole healing area determination and node relocation.

## 6. REFERENCES

- 1) Taeyoung Kim,<sup>1</sup> Minhan Shon,<sup>1</sup> Mihui Kim,<sup>2</sup> Dongsoo S. Kim,<sup>3</sup> and Hyunseung Choo<sup>1</sup> “Anchor-Node-Based Distributed Localization with Error Correction in Wireless Sensor Networks”, 2012
- 2) “Accurate Anchor-Free Node Localization in Wireless Sensor Networks” Adel Youssef, Ashok Agrawala and Mohamed Younis.
- 3) “Movement-Assisted Sensor Deployment” Guiling Wang, Guohong Cao, and Tom La Porta.

- 4) “Movement Control Algorithms for Realization of Fault-Tolerant Ad Hoc Robot Networks” Prithwirh Basu and Jason Redi.
- 5) K. Akkaya, F. Senel, A. Thimmapuram, and S. Uludag, “Distributed recovery from network partitioning in movable sensor/actor networks via controlled mobility,” IEEE Trans. Comput., vol. 59, no. 2, pp. 258–271, Feb. 2010.
- 6) K. Akkaya and M. Younis, “COLA: A coverage and latency aware actor placement for wireless sensor and actor networks,” in Proc. IEEE VTC, Montreal, QC, Canada, Sep. 2006, pp. 1–5.
- 7) A. Youssef, A. Agrawala, and M. Younis, “Accurate anchor-free localization in wireless sensor networks,” in Proc. 1st IEEE Workshop Inf. Assurance Wireless Sensor Netw., Phoenix, AZ, Apr. 2005.
- 8) S. Vemulapalli and K. Akkaya, “Mobility-based self route recovery from multiple node failures in mobile sensor networks,” in Proc. 10th IEEE Int. Workshop WLN, Denver, CO, Oct. 2010.
- 9) S. Yang, M. Li, and J. Wu, “Scan-based movement-assisted sensor deployment methods in wireless sensor networks,” IEEE Trans. Parallel Distrib. Syst., vol. 18, no. 8, pp. 1108–1121, Aug. 2007.
- 10) G. Wang, G. Cao, and T. La Porta, “Movement-assisted sensor deployment,” IEEE Trans. Mobile Comput., vol. 5, no. 6, pp. 640–652, Jun. 2006.



# Isolated Arabic Handwritten Character Recognition Using Linear Correlation

Abd Elhafeez Hamid    Mariam M. Musa    Mona M. Osman    Moahib M. Alamien    Marwa M. Elsaied  
Dept. of CS            Dept. of CS            Dept. of CS            Dept. of CS            Dept. of CS  
Faculty of CS & IT    Faculty of CS & IT    Faculty of CS & IT    Faculty of CS & IT    Faculty of CS & IT  
Kassala University    Kassala University    Kassala University    Kassala University    Kassala University  
Kassala-sudan            Kassala-sudan            Kassala-sudan            Kassala-sudan            Kassala-sudan

**Abstract:** Handwriting recognition systems have emerged and evolved significantly, especially in English language, but for the Arabic language, such systems did not find that sufficient attention in comparison to other languages. Therefore, the aim of this paper to highlight the Optical Character Recognition using linear correlation algorithm in two dimensions and then the programs can to identify discrete Arabic letters application started manually, the program has been successfully applied.

**Keywords:** Optical Character Recognition; Handwriting; Image Processing; Pattern Recognition; off-line handwriting recognition;

## 1. INTRODUCTION

We ask that authors follow some simple guidelines. This Pattern recognition is the scientific discipline whose goal is the classification of objects into number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified.

The handwriting recognition refers to the identification of written characters. Handwriting recognition has been become a very important and useful research area in recent years for the ease of access of many applications. [1]

There are two types of handwriting recognition: off-line recognition and on-line recognition. Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting. Off-line handwriting recognition is comparatively difficult, as different people have different handwriting styles. On-line handwriting recognition involves the automatic conversion of text as it is written on a special digitizer or PDA, where a sensor picks up the pen-tip movements as well as pen-up/pen-down switching. This kind of data is known as digital ink and can be regarded as a digital representation of handwriting. The obtained signal is converted into letter codes which are usable within computer and text-processing applications.

## 2. ARABIC CHARACTERS

All Arabic characters are used in writing many languages not only in Arabic countries, but for Urdu and Farsi and other languages in countries where Islam is the principal religion (e.g., Iran, Pakistan, and Malaysia). [4] The special characteristics of Arabic written words and characters do not allow the direct application of algorithms for other languages. See figure 1.

|      |       |       |       |       |       |       |
|------|-------|-------|-------|-------|-------|-------|
| خ    | ح     | ج     | ث     | ت     | ب     | أ     |
| Xaa' | H'aa' | Jeem  | Thaa' | Taa'  | Baa'  | 'Alif |
| ص    | ش     | س     | ز     | ر     | ذ     | د     |
| Saad | Sheen | Seen  | Zaay  | Raa'  | Thaal | Daal  |
| ق    | ف     | غ     | ع     | ظ     | ط     | ض     |
| Qaaf | Faa'  | Ghayn | 'Ayn  | Thaa' | Taa'  | Daad  |
| ي    | و     | هـ    | ن     | م     | ل     | ك     |
| Yaa' | Waw   | Haa'  | Noon  | Meem  | Laam  | Kaaf  |

Figure. 1 Arabic letters

Arabic's Letters characteristics are:

- Arabic is a cursive type language written from right to left.
- Arabic has 28 basic characters. Each character has 2-4 forms depending on its position within the word.
- Many letters of the Arabic alphabet have dots, above or below the character body, and some letters have a Hamza (zigzag shape) and dilation.
- Overlapping characters: some Arabic's characters become over each other horizontally when they connected with each other. [4]

**Table1. Arabic characters and their shapes at different positions in the word**

| Letter | Stand-alone | Initial | Middle | Final | Other shapes |
|--------|-------------|---------|--------|-------|--------------|
| Alef   | ا           |         |        | آ     | آي           |
| Ba'    | ب           | ب       | ب      | ب     |              |
| Ta'    | ت           | ت       | ت      | ت     | ة            |
| Tha'   | ث           | ث       | ث      | ث     |              |
| Jeem   | ج           | ج       | ج      | ج     |              |
| H'a'   | ح           | ح       | ح      | ح     |              |
| Kha'   | خ           | خ       | خ      | خ     |              |
| Dal    | د           |         |        | د     |              |
| Thal   | ذ           |         |        | ذ     |              |
| Ra'    | ر           |         |        | ر     |              |
| Zai    | ز           |         |        | ز     |              |
| Seen   | س           | س       | س      | س     |              |
| Sheen  | ش           | ش       | ش      | ش     |              |
| Sad    | ص           | ص       | ص      | ص     |              |
| Dhad   | ض           | ض       | ض      | ض     |              |
| Tta    | ط           | ط       | ط      | ط     |              |
| Dha'   | ظ           | ظ       | ظ      | ظ     |              |
| Ain    | ع           | ع       | ع      | ع     |              |
| Gahin  | غ           | غ       | غ      | غ     |              |
| Fa'    | ف           | ف       | ف      | ف     |              |
| Qaf    | ق           | ق       | ق      | ق     |              |
| Kaf    | ك           | ك       | ك      | ك     |              |
| Lam    | ل           | ل       | ل      | ل     |              |
| Mecm   | م           | م       | م      | م     |              |
| Noon   | ن           | ن       | ن      | ن     |              |
| Ha'    | ه           | ه       | ه      | ه     |              |
| Waw    | و           |         |        | و     |              |
| Ya'    | ي           | ي       | ي      | ي     |              |

### 3. OCR system

OCR systems consist of five major stages:

#### 3.1 Image acquisition

Is the first step in the algorithm, the system takes the original image that needed to read, from the scanner or from computer storage, this process can be represented in follow chart as shown in Figure 2:

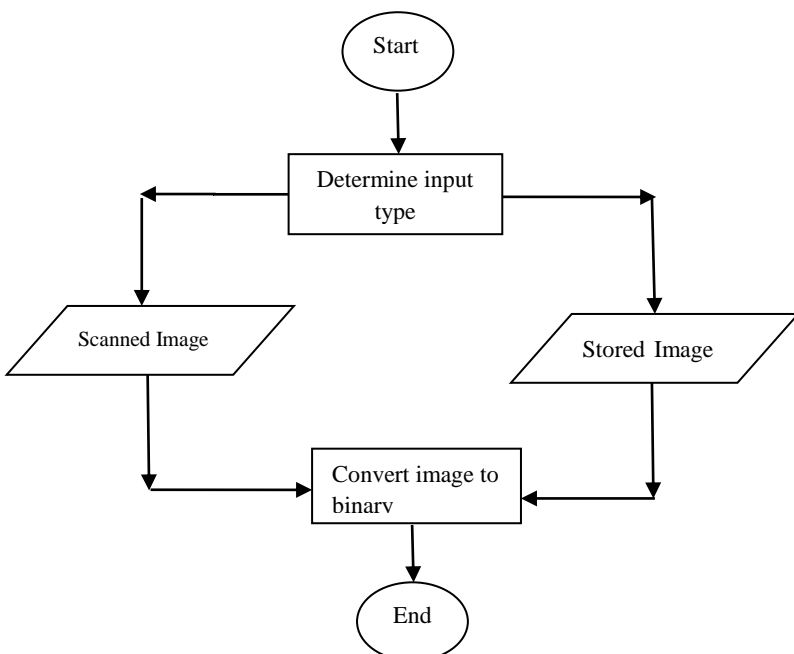


Figure. 2 Image acquisition algorithm

### 3.2 Pre-processing

The aim of preprocessing stage is the removal of all elements in the word image that are not useful for recognition process. It includes:

#### 3.2.1 Binarization:

Binary images are the simplest type of images that take one of two values, typically black and white, or '0' and '1'. A binary image is referred to as a 1 bit/pixel image because it takes only 1 binary digit to represent each pixel. This type of image is most frequently used in computer vision application where only information required for the task is general shape, or outline, information. [3]

#### 3.2.2 Smoothing:

Filling gaps and eliminating superfluous points of the contour image.

#### 3.2.3 Cleaning:

Removing noise that could not be eliminated by smoothing.

#### 3.2.4

Determine the size and refine the distortions of the image and their impurities which may be associated with it.

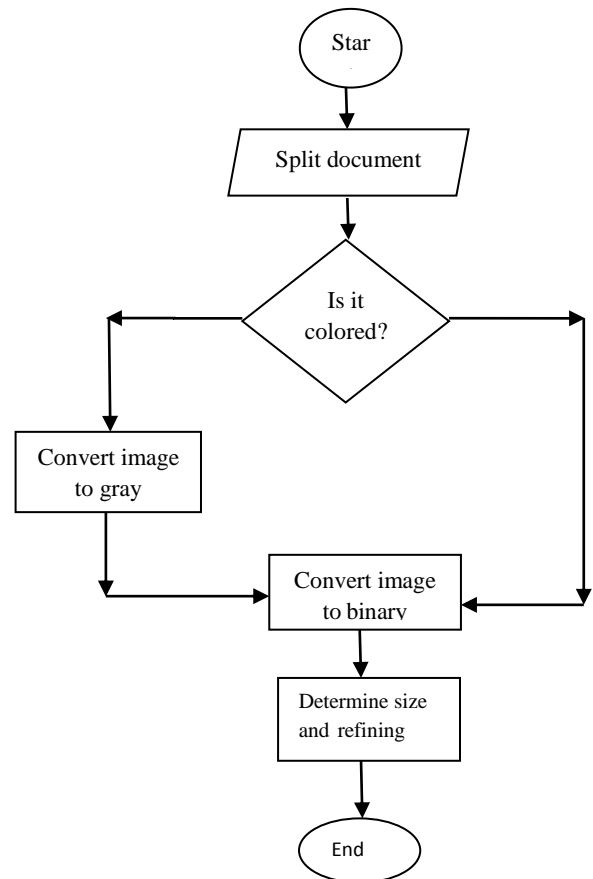


Figure. 3 Pre-process algorithm

### 3.3 Segmentation

3.3.1 First the document is divided into lines by using histogram, then calculating the number of dots in each horizontal pointed row.

3.3.2 Dividing the lines of the document into Characters, according to the shape of the Character, based on rules and information that owned by the system.

3.3.3 Extracting the features by collecting the dots in each row separately, and also to the columns, then studying and analyzing the characteristics such as Character height and width, in preparation to identify the Crafts.



Figure. 4 image segmentation into lines and characters

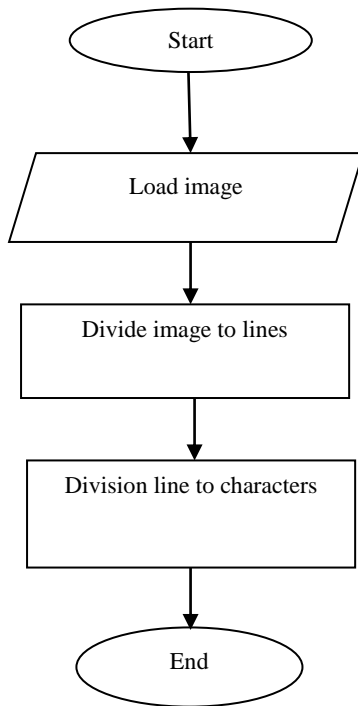


Figure. 5 Segmentation algorithm

### 3.4 Feature Extraction

This process extracts the features of the characters that are most relevant for classifying at recognition stage. This is an important stage as it can help avoid misclassification, thus increasing recognition rate.

In feature extraction stage each character is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

### 3.5 Classification

Classification is a variety of ways including linear correlation algorithm the linear correlation matrix phase is compared to the characters to be identified with matrices stored in the data base and are selected after comparison the largest correlation value is determined by the appropriate letter calculated the correlation between the value of the following formula:

$$x = \sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})$$

$$y = \sqrt{\left( \sum_m \sum_n (A_{mn} - \bar{A})^2 \right) \left( \sum_m \sum_n (B_{mn} - \bar{B})^2 \right)}$$

$$r = \frac{x}{y}$$

Where:

$r$  = correlation value.

$A$  = initial matrix (for character want to identify it).

$B$  = template matrix (for stored character).

$\bar{A}$  = mean of the initial matrix (character want to identify it).

$\bar{B}$  = mean of the template matrix (for stored character).

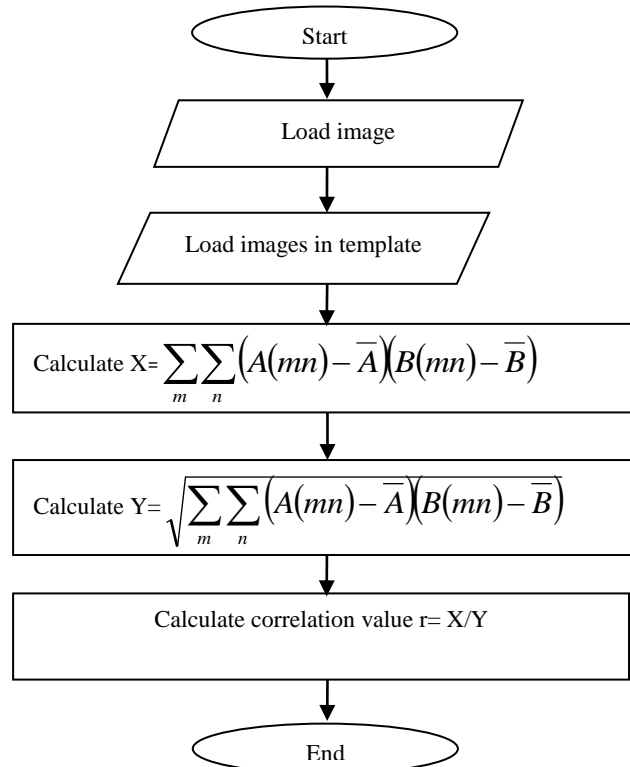


Figure. 6 linear correlation algorithm

At the end, the work of this algorithm can be summed up in the following chart:

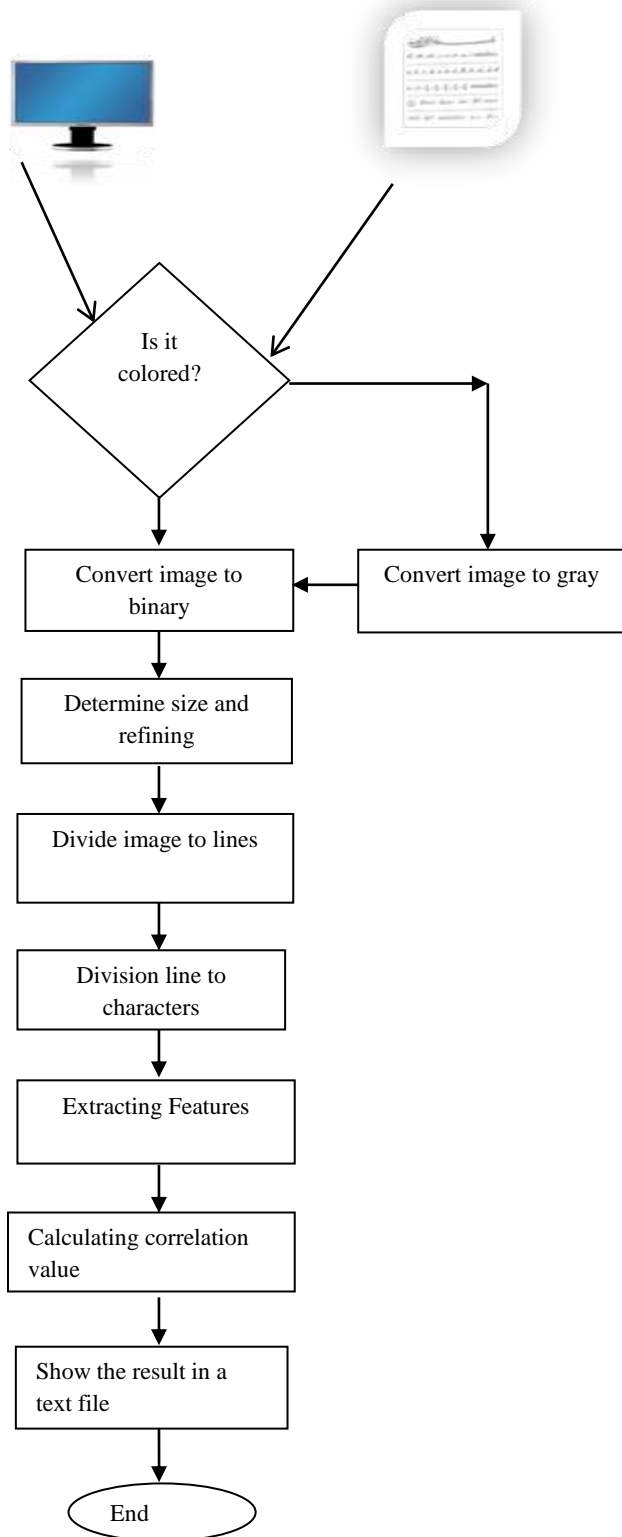


Figure. 7 : completely algorithm

### 3.6 Evaluating The Results

After the application of the proposed algorithm on the number of images we calculated the correlation factor and the ratio Signal to Noise Peak Signal-to-Noise Ratio (PSNR) between the input image and the resulting images, and is the PSNR account the following law:

$$PSNR = 10 * \log (255 * 255 / MSE) / \log (10)$$

The law used to calculate the mean square error (MSE) is:

$$MSE = \text{sum} (\text{sum} (\text{error} * \text{error})) / (M * N)$$

The results were as follows:

TABLE2. Recognition performances

| no | Recognition performances of the PSNR |         |
|----|--------------------------------------|---------|
|    | Character                            | PSNR    |
| 1  | أ                                    | 32.0951 |
| 2  | ب                                    | 29.8934 |
| 3  | ت                                    | 33.0398 |
| 4  | ث                                    | 36.6835 |
| 5  | ج                                    | 36.6684 |
| 6  | ح                                    | 35.8495 |
| 7  | خ                                    | 35.8950 |
| 8  | د                                    | 30.8357 |
| 9  | ذ                                    | 33.9198 |
| 10 | ر                                    | 32.8113 |
| 11 | ز                                    | 35.2482 |
| 12 | س                                    | 30.5137 |
| 13 | ش                                    | 35.8866 |
| 14 | ص                                    | 36.2440 |
| 15 | ض                                    | 30.5039 |
| 16 | ط                                    | 27.1389 |
| 17 | ظ                                    | 28.4566 |
| 18 | ع                                    | 35.1700 |
| 19 | غ                                    | 30.5559 |
| 20 | ف                                    | 34.8221 |
| 21 | ق                                    | 31.1804 |
| 22 | ك                                    | 37.2377 |



| no | Recognition performances of the PSNR |         |
|----|--------------------------------------|---------|
|    | Character                            | PSNR    |
| 23 | ل                                    | 33.2289 |
| 24 | م                                    | 28.6027 |
| 25 | ن                                    | 36.1048 |
| 26 | هـ                                   | 30.5709 |
| 27 | و                                    | 36.7898 |
| 28 | ي                                    | 34.7908 |

#### 4. REFERENCES

- [1] **Abdelhafeez hamid**, G. Zen Alabdeen, A. Mansour, "Arabic Numerals Recognition Using Linear Correlation Algorithm in Two Dimensions", International Journal of Computer Applications Technology and Research Volume 4– Issue 1, 01 - 06, 2014.
- [2] K. Kaur<sup>1</sup>, N. K. Garg, "A Survey on Process of Isolated Character Recognition", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2014.
- [3] Haithem Al-Ani, Ban N, Dhannoon and Huda M. Abass, "Printed Arabic Character Recognition using Neural Network", Journal of Emerging Trends in Computing and Information Sciences, Vol. 5, No. 1 January 2014.
- [4] Hasan Al-Rashaideh, "Preprocessing phase for Arabic Word Handwritten Recognition", Information Transmition In Computer Networks, 2006.
- [5] Shatha M. Noor, "Isolated Multifont Arabic Character Recognition Using Fourier Descriptors", Journal of Al-Nahrain University Vol.16 (1), March, 2013, pp.176-182.
- [6] Ahmed M. Shaffie and Galal A. Elkobrosy, "A Fast Recognition System for Isolated Printed Characters Using Center of Gravity and Principal Axis", Applied Mathematics, 2013, 4, 1313-1319.
- [7] M. A. Abdullah, L. Al-Harigy, and H. Al-Fraidi, "Off-Line Arabic Handwriting Character Recognition Using Word Segmentation", Journal of computing, Vol. 4, Issue 3, March 2012, ISSN 2151-9617.
- [8] F. Parrweej, "An Empirical Evaluation of Off-line Arabic Handwriting And Printed Characters Recognition System", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 1, November 2012.
- [9] B. Alijla and K. Kwaik, "Online Isolated Arabic Handwritten Character Recognition Using Neural Network", The International Arab Journal of Information Technology, Vol. 9, No. 4, July 2012.
- [10] A. Pal and D. Singh, "Handwritten English Character Recognition Using Neural Network", International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 141-144.
- R. Kala, H. Vazirani, A. Shukla and R. Tiwari, "Offline Handwriting Recognition using Genetic algorithm", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010 .
- [11] M. Al-Alaoui, M. Amin, Z. Abou Chahine, and E. Yaacoub, "A New Approach for Arabic Offline Handwriting Recognition", IEEE Multidisciplinary Engineering Education Magazine, Vol. 4, No. 3, Semptember 2009.
- [12] H. Almohri, J. S. Gray, H. Alnajjar, "A Real-time DSP-Based Optical Character Recognition System for Isolated Arabic characters using the TI TMS320C6416T", IAJC-IJME International Conference, 2008, ISBN 978-1-60643-379-9.
- [13] Mansoor Al-A'ali and Jamil Ahmad, "Optical Character Recognition System for Arabic Text Using Curvilinear Multi-Directional Approach", Journal of Computer Science 3 (7): 549-555, 2007 ISSN 1549-3636.

# Analytical Comparison of Mathematical Modeling in the Diagnostic Expert Systems

ALI AMOOJI  
 Payam-e-noor University  
 Tehran, Iran

**Abstract:** In the article diagnosis problem as the main issue has been investigated and as instance of diagnostics issues, medical diagnosis has been researched. Usable and unusable mathematical principles in the expert systems have been analyzed in the next step. Bayesian formula for calculate of occurrence probability of a hypothesis (assumption) has been surveyed. As well as utilization of this formula for calculating the probability of a disease is suggested with justifiable reasons. To show the function of these mathematical principles in reasoning cycle of each expert system, causal-covering method has been investigated. This method discloses the function of inference engine in expert systems. The determined diagnosis is actually comparative cognition.

**Keywords:** diagnosis issue, inference engine, mathematical modelling, Bayesian method, causal-covering method.

## 1. INTRODUCTION

Nowadays expert systems are a part of artificial intelligence so that they are very useful and efficient. Expert systems rely on proficiency of specialists and accuracy of computer, improve the quality of decisions. On the other hand diagnosis-based issues are used in the design of inference engine of expert systems. Initially it is necessary to explain briefly about the diagnosis issue and the inference engine [9].

Ultimate goal of each diagnostic issue is satisfaction from the obtained results. Therefore, the experts use the goal oriented methods for correcting diagnosis (Goals driven methods) [4]. Also inference engines are structured controllers that using the rules for obtain the results. These controllers are known as the decision-making engines. Their main role is use and process of exist information in the knowledge bases and solve the certain issue [1]. Generally the functions of inference engine in each cycle are selection, comparison and elimination of the controversial cases stages. In each stage three source of knowledge are used: working memory, situation memory and knowledge base [3].

## 2. ANALYZING THE MATHEMATICAL PRINCIPLES IN THE DIAGNOSIS ISSUES.

The ultimate goal of each diagnostic issue is satisfaction from the obtained results. Therefore, the experts use the goal oriented methods for correcting diagnosis (Goals driven methods) [8]. Two mathematical rules are used in diagnostic issues. First rule was expressed as:

$$(A \equiv \text{TRUE}) \wedge (A \rightarrow P) \Rightarrow \forall a \in A (P \equiv \text{TRUE})$$

If A is true, and P is outcome of A, and a is sample of A, then P must be true [7].

The second mathematical principle that used in diagnostic issues is like this: B is the outcome of A, if B contradictory is false and A and a is sample of A then A is false:

$$(A \rightarrow B) \wedge \sim B \Rightarrow \forall a \in A, A = (\sim A)$$

These rules are used in the logic of inference engine. But there are some of the rules in the inference engine that are in

conflict with mathematical rules. For example, there is a rule that applied in expert systems but is not valid as mathematical principle. This is called deny rule [6].

$$(A \rightarrow B) \wedge \sim A \Rightarrow \forall b \in B, B = (\sim B)$$

Secondary mathematical law that is valid in the world of mathematics, but it cannot be used to inference engine of expert systems, called nominally law:

$$\begin{aligned} \forall x \in (B \subseteq A) : F(x) \rightarrow G(x) \\ \forall x \in A : F(x) \rightarrow G(x) \end{aligned}$$

According to this mathematical principle, if for all members of A majority,  $f(x) \rightarrow G(x)$  relation is valid and B is subset of A, so  $f(x) \rightarrow G(x)$  relation is valid for all members of B. This mathematical principle does not meet the diagnosis expert systems, because for current situation of A majority,  $f(x) \rightarrow G(x)$  relation is valid but in the same situation for all members of B majority,  $f(x) \rightarrow G(x)$  relation is not valid.

Uncertainty nature of diagnosis issues leads us to one of the probability principles. Since the data of diagnosis systems are uncertain accordingly the uncertainty analysis is necessary [5]. For implementation of uncertainly, the Bayesian inference probabilistic method can be suitable. Bayesian method reflects the uncertainty in a range of assumptions and presumptions. Thus selected problem in the expert systems are placed in the hypothesis.

Occurrence probability of the hypothesis can be found. Bayesian method determines the value of initial occurrence probability of hypothesis [7].

Of course, this evaluation is not final result, it just updated in the subsequent steps. Value of the specified assumptions is obtained via following formula:

$$P(D_i | E) = [p(E | D_i) * p(D_i)] / p(E)$$

$$P(D_i | E) = \frac{[p(E | D_i) * p(D_i)]}{[\sum_{j=1}^i p(E | D_j) * p(D_j)]}$$

**Di:** i-nd problem  
**E:** symptom and sign of problem  
**p(E):** initial occurrence probability of problem  
**P(Di | E):** the E symptom is valid and Di is the i-nd problem  
**p(E | Di):** Di (i-nd) problem is valid and occurrence probability of E symptom  
**n:** number of problem (hypothesis) [8].

For matching the above mathematical rules with the medical diagnosis case first of all deny rule be surveyed:

$$(A \rightarrow B) \wedge \sim A \Rightarrow \forall b \in B, B = (\sim B)$$

The rule can be used in expert systems but is not valid as mathematical principle. Because for example to the medical diagnosis, we do not cancel the surgery for a patient so that the patient does not need any surgery, perhaps the problem of return to conscious can cancel the surgery.

### 3. SCRUTINY OF MEDICAL DIAGNOSIS AS A SAMPLE OF DIAGNOSIS ISSUES.

In most cases, doctors to correct diagnosis, tend to use the agents that are based on Goals and Driven methods. Therefore, the doctors trying to treatment of the patient's disease based on their diagnoses. Certainly, the ultimate goal of this process is the patient's satisfaction. That is why, the problem to be resolved expediently. The rapid development of medical science and the nature of uncertainty of medical knowledge increased the complexity of diagnostic issue. Physicians are doing process of diagnose the disease in the following order:

1. All diseases symptoms are determined;
2. All illnesses that their signs are observed in the patient is intended;
3. Probabilities of occurrence of estimated diseases are rated;
4. Selection of diseases that based on final rating their occurrence probabilities are high. Additional questionnaires filled, additional examinations, inspections and analyzes are offered;
5. If any diseases are not properly selected subsequent high probability disease is tested;
6. If appearance of new symptoms on the patient or results of pathology analysis can come doctor to a new disease, it is proposed to repeat algorithm for new disease that founded;
7. If the diagnosis of a disease was finalized, doctor can prescribe the match drugs and other suggestions.

As well as implement of Bayesian formula to calculate the occurrence probability of disease will be as follow:

**Di:** i-nd disease  
**E:** a symptom and sign of disease  
**P(E):** initial occurrence probability of disease  
**P(Di | E):** Di is the i-nd disease, on the condition that the E symptom is exists to the patient  
**P(E | Di):** occurrence probability of E symptom, on the condition that Di (i-nd) disease is exists  
**n:** number of diseases.

### 4. ANALYSIS OF INFERENCE ENGINE IN THE MEDICAL DIAGNOSIS ISSUE.

First of all, pseudo code of the causality map is writing:

While new-solution cover (complaint -> hypothesis) do  
 Differential: = hypothesis add differential;  
 End while  
 Repeat

Select (differential -> hypothesis);  
 Specify (hypothesis -> observable);  
 Obtain (observable -> finding);  
 Evidence: = finding, add evidence;  
 For each hypothesis in differential do  
 Verify (hypothesis + evidence -> result);  
 If result = false, then differential: = differential subtract hypothesis  
 Until length differential = <= 1 or "no observables left"  
 Faults: = hypothesis;

Here we need to explain the tools that give us results. Below is review of the five functions.

**Cover:** this inference engine selects the hypotheses that have certain symptoms and signs. In this stage many hypotheses can be selected. This hypotheses separately kept in the different place.

**Select:** this inference engine chooses the one of the selected hypotheses that its occurrence probability is high.

**Specify:** this inference engine compares the selected hypothesis with observed signs. This observed conformity may lead to another hypotheses. Of course it is also possible to exclude the occurrence of cases.

**Obtain:** This is a function that examines the selected assumptions (problem), the reasons adapt and assess their value.

**Verify:** this inference engine directs the founded hypothesis to the output.

The later four functions are located in the loop. All of selected hypotheses test to the loop. This cycle continues until there is one of the assumptions then the cycle will be canceled. This process is performed via select inference engine [9].

In any case, task-oriented method, could lead the result to three situations:

- The result is empty: the type of disease is not found. The evidence does not match together all diseases that are known in system.
- When Disease has been found, determining the diagnosis accurately. Usually this is an ideal result.
- Several diseases are likely occurred: system keeps these assumptions, examines and inspects the patient or suggests the additional analysis.

### 5. CONCLUSION

The process of diagnosis is the lack of the knowledge that reveals the evidence to the different classes. Diagnosis Looks to the classes of the domain for evidence, signs and indicators. For solving diagnostic issue, Bayesian formula is useful for all mathematical models. The advantages of this model can be as follow:

- Through the Bayesian method creating of causal map is possible.
- Using this causality map with Bayesian calculation leads us to trust calculation.

Often creating and testing strategies are used in the inference engine of diagnostic expert systems. Diagnostic task divided

into five sub functions (four inference engines and one transfer function).

## 6. REFERENCES

- [1] Bruce P. Car Wars: Birth of an Expert System // AI Expert, 1998, v.12, No 4, 60-64
- [2] Kərimov S. Q., Şükürlü S. F., Oftalmologiya sahəsi üzrə ekspert sistemində çıxarış mexanizminin təşkili // AMEA News, 2005, No 2, 80-85
- [3] Kərimov S.Q., 2004, Ekspert sistemləri.
- [4] Robinson J. A Machine-Oriented Logic Based on the Resolution Principle // J.ACM, 2007, v.12, No 1, 23-41
- [5] Savage L. 2007, The Foundation of Statistics.
- [6] Shafer G. Mathematical theory of Evidence// Princeton University Press, 2006, v.12, No 8, 24-35
- [7] Smith M. 2008, Neural Networks for Statistical Modeling.
- [8] Zadeh A. Review of book A Mathematical theory of Evidence // AI Magazine, 1994, v.12, No 4, 81-83
- [9] Defog R., Bemus B., and Volker H. The Common KADS organization model: content, usage and computer support // Expert Systems With Applications, 1996, v. 2, No 1, 29-40

# Predicting the Impact of a Technology for Instantly Verifying the Licenses of Vehicles/Drivers in Ghana

Abeo Timothy Apasiba  
School of Applied Science  
Dept. of Computer Science  
Tamale Polytechnic  
Tamale, Ghana

Ernest D. Ganaa  
School of Applied Science &  
Technology  
Department of ICT  
Wa Polytechnic  
Wa, Ghana

Zakaria Abukari  
School of Applied Science  
Dept. of Computer Science  
Tamale Polytechnic  
Tamale, Ghana

---

**Abstract:** Vehicle and driver license registration is one of the many commodities that contribute significantly to the revenue generation capabilities of Ghana. However according to the investigation carried out for this research, 86.3 percent of respondents attest to the existence of fake vehicle and driver licenses. Thus, the government of Ghana loses a lot when it comes to generating revenue from Driver and Vehicle Licensing Authority. This research found out a yearly revenue generation capacity from the Tamale office of the Driver and Vehicle Licensing Authority of about GHS1, 440,000 from only vehicle registration, amidst this massive evasion by drivers and vehicle owners.

This research unearthed the major factors motivating the menace of non-licensing of vehicles and drivers; it as well exposed the methods by which vehicles are being stolen from their owners, pointing out reasons for vehicle owners' reluctance to report for the recovery of these stolen vehicles. The technology employed for license verification by the security services was not left out.

**Keywords:** Vehicle Theft; DVLA Ghana; License Verification; Ghana Revenue; Driver; Technology;

---

## 1. INTRODUCTION

The massive evasion of licenses for drivers and vehicles results in a number of related crimes. This is seen to be on the ascendancy with an almost daily occurrence in the country and the world at large. According to the Ghana Statistical Service report of May 2010, titled "*Victimization survey in Ghana (2009)*" reported cases of vehicles and related crimes ranged from 5.9% for the crime of "theft from car" to 83.3% for "theft of car". With the other related crimes falling in between the two percentages.

Also, statistics show that road accidents kill an average of four people a day in Ghana. In the year 2005, there existed 16% increase in road accidents as compared to the preceding year, (Global Road Safety Partnership 2012). As appeared in [www.ghanaweb.com](http://www.ghanaweb.com), between 2007 and 2010 the Motto Traffic and Transport Department (MTTD), in its report noted at least 6,000 people had died due to road accidents with an additional 40,000 injured within the same interval. This takes away valuable human resources which could have contributed immensely to the development of the country. Also, funds which should have been channeled into developmental projects are diverted into medical expenses of victims, damage to vehicles and insurance cost among others. According to Ghana News Agency (2008) it is estimated that Ghana loses an amount of GHS165, 000 annually which represents 1.6 percent of Gross Domestic Product (GDP) through road accidents.

This evasion of licenses causes the nation a great loss of revenue. For example in the year 2012 alone, the Ashanti Regional Office of the Driver and Vehicle Licensing Authority (DVLA) generated GHS4,596,379.30 as revenue from vehicle registration, renewal of

road worthy certificates, licenses and sale of highway codes to motorists ([ghanabusinessnews.com](http://ghanabusinessnews.com) 2012). Sums of this magnitude can easily be lost by the state if the currently high levels of evasions are allowed to prevail

## 2. OBJECTIVES OF THE RESEARCH

This research seeks to establish the impact of evasion of vehicle licensing to the development of the country.

The specific objectives;

- to forecast the cost implications with the introduction of a license verification technology
- to uncover the reasons behind evading vehicle/driver license registration and or fake licensing
- to reveal the modes of vehicle theft in the country to aid the intelligence of the security services

## 3. EXISTING SYSTEMS OVERVIEW

The importance of the study of relevant and related literature on research work cannot be overemphasized. This eliminates the trouble of having to go through the same or similar mistakes made by previous researchers. It will as well give much insight as to the relevance of the research to be conducted. According to Chaplin as cited in Cisse (2006), it is short-sightedness and a waste of time to plunge at once into research without first taking a look at what has been done.



In the area of vehicle security, a little talk about the better, it has been left solely to the vehicle owners' themselves to manage, with the DVLA not making extra efforts aside keeping a database of registered vehicles. In this regard Abeo et al (2015) saw the need to develop a technology to assist in the instant verification of driver and vehicle licenses by the security services. From their research they found out that vehicle theft occurs daily with an endorsement from 90 percent of respondents. Their research gave birth to a GSM/SMS technology that when integrated into the driver and vehicle licensing authority's database, could instantly give a reply with vital information relating to the vehicle and the owner when a security personnel sends a mobile text message to it.

There is no doubting the extensiveness of this research. On the other hand, they could not delve into the revenue generation ability which the implementation of this technology would have supported.

*Bracha* (2008) saw the need to invent a system for verifying Driver's insurance coverage. In his system, a database of Driver's License ("DL") number identifying a motorist is maintained. A request containing a DL is used to fetch an indexing DL in the database and retrieving the DL-indexed motorist record, which they discovered as the only proof of insurance.

As could be noted, *Bracha's* (2008) concentration is narrowed to only insurance coverage monitoring. They could not move further to ascertain the reasons for possible low insurance coverage.

Taking vehicle monitoring into a different angle, *Zlojutro* (2010) designed a vehicle monitoring and traffic enforcement system in which a wireless communication device is linked to motor vehicles. Data about the identification of the vehicle could be sent to a database which contained up to date information concerning insurance law compliance, vehicle registration and the like. When a violation is detected on any, the system sends real time information to the law enforcement agency and also to a nearby law enforcement vehicle using GPS (Geographical Positioning System) and GIS (Geographic Information System) technologies. The information provided by the system does not only show the vehicle being tracked but in addition an area map showing roads, traffic situations as well as the location of other emergency vehicles for law enforcers to be able to act promptly.

A problem foreseen in this system is that, there will be a failure in the final execution. Ideally, the system is expected to work automatically without a lot of human intervention; this might not apply very well in our society where the law enforcement agencies like the police service most at times expects the victims to keep making follow-ups on them after reporting their cases. Also, this research could not establish reasons for people not reporting to the security services when their vehicles are stolen.

Adding more functionality to check post or Toll base, *Mahesh* (2012) developed a smart logic system which helps to detect stolen vehicles at the check post or Toll base. This technology

seems to be the first of its kind in its target area. In this technology, a Unique Identification (UID) code is associated with every vehicle. Each and every vehicle is embedded with a silicon chip which contains that unique number pre-installed. This UID chip is strategically placed in the engine of the vehicle to avoid easy manipulation by criminals. A database containing vehicle codes is also stored and maintained at the check post or Toll base. It is then made possible for any vehicles' passing through the check post/Toll, Radio Frequency (RF) tag (which has been pre-installed in the vehicle) to be decoded by a 125kHz frequency generated by an RF passive vehicle reader. The code is then checked with that of the database, if the stolen UID code is not found then the security gate gets open and the vehicle is able to proceed. On the contrary where the stolen UID code finds a match in the database then the security gate remains closed and an alarm is then triggered ON automatically. This draws the attention of the nearby authorities and finally, the vehicle is apprehended by the security apparatus.

Having gone through this research, there is no doubt that if this system is ideally implemented in an ideal society it would have been a perfect system for tracking fake vehicle license. However, it comes with a lot of limitations as follows; the gates at the check post or Toll base have to function automatically without human intervention, which will be difficult to achieve in our part of the world, there is non-existent of even manual gates. The cost of implementing such a system will be quiet huge, since you will need to deploy experts to various check post culminating with installation of databases in all such areas. These systems are all electronic and thus there must be constant supply of electricity or else gates at check post or Toll base could seize to open or vice versa. The study failed to do a cost-benefit-analysis for the implementation of this system.

*Shihab et al* (2012) asserted that, a lot of efforts are being made by local and international companies towards producing car security systems, but the number of car theft cases kept on increasing beyond expectation. It added, "The thieves are developing and inventing smarter and stronger capabilities and stealing techniques which require very powerful security systems. *Shihab et al* (2012) realized most of the security systems made use of traditional alerts which do not support multimedia and mobile facilities. Also the few existing advance car security systems are very expensive and are commonly used in expensive cars. Thus they proposed an "effective Car Monitoring and Tracking Model" (CMTM) which tried addressing the above problems. In their research, they introduced a powerful security model that sends SMS and MMS to the owner or security organization for a quick response. In their system, a database containing information about cars and their owners is maintained. The system made provisions for the police or security authorities to track the car using GPS systems which is linked to Google Earth as well as other mapping software.

Still delving more into the research of car security, *Abdullah* (2011) noted that; the existing vehicle alarm systems are of no match to the well-equipped thieves. There may be so many car alarms that are too sensitive, while the rest can withstand a major

earthquake without a single beep. Also, it has limited capability to interact with its owner.

In addition, the conventional alarm systems did not help in any way in assisting the recovery of a stolen vehicle. While the GPS systems do, there are situation where the GPS system encounters a deficiency such as at underpasses and indoor parking. In view of these challenges they came out with a “Mobile Controlled Car Security System” (MCCS). This mobile controlled system is capable of providing a two way effective communications between the alarm system and the owner of the car. In this system when an intrusion is detected, by the help of the mobile transceiver a call is quickly made to the owner of the car immediately. In addition, there is a feature that makes it possible for the owner to remotely control the car through a phone call. The mobile phone transceiver is detected by the Mobile Switch Center (MSC) of the GSM network as long as the phone is turned on. The user can then request for the location of the car from the MSC.

In the heat of finding solutions to the problem of vehicle security, one is likely to gross over the cost of implementation of such a system. Khondker et al (2009) noted the expensive nature of most of the existing live tracking systems which employ SMS for communication to the server. Thus, they proposed a “Cost Effective GPS-GPRS Based Object Tracking System” which solves the underlining problem was carried out. In this system a MYSQL database is maintained which is interfaced with a web server, with embedded Google Maps. The target vehicle location is read by the system using GPS; this data is then sent using the services of GPRS from the GSM network to the web server. The information is finally stored into the database for both past and live tracking. The past and present recorded positions are then available through the internet on Google Maps for the user to access. It was concluded that this system was very useful for situations of car theft, for parents to watch and monitor their adolescent drivers as well as tracking both pets and humans alike.

## 4. RESEARCH APPROACH

Taking into consideration the nature of the research, both qualitative and quantitative approaches were adopted; so as to be able to have answers to questions interviewee may feel reluctant to give responses to, thus the qualitative approach.

### 4.1 Sample Population and Techniques

The study considered a 152 population size; distributed as 7 staff from the Driver and Vehicle Licensing Authority, 45 officials from MTTD coming from the Ghana Police service, and both private and commercial vehicle owners/drivers being 100, all from the Tamale metropolis. The choice was carefully made since they were the core actors who were directly linked to the objectives of the study, and these were the group of people who were interviewed for the data collection.

The study generally made use of purposive, quota and convenience sampling techniques. When it came to getting the technical people from DVLA and MTTD, the Manager and Commander from DVLA and MTTD respectively and the leaders

from the various unions of the commercial vehicle stations, it was the purposive sampling technique which was employed.

In dealing with the commercial and private vehicle owners, drivers and other official of the MTTD, there was a combined quota and convenience sampling techniques which were used. This was based on the availability of the respondent at that particular time while being mindful of the sample allocation to the various groups. This was how sampling was applied to technically extract balanced respondents.

## 4.2 Methodology

This study made use of techniques in the form of Interviews and questionnaires in the collection of the required data and information necessary for answering the objectives of the research. In reality the qualitative data which was gathered from the verbal interviews and observations helped greatly to re-structuring the questionnaire.

In this study, the research findings and analysis were presented in accordance to the research objectives. The general objective of this study seeks to establish the impact of evasion of vehicle licensing to the development of the country. The data for this study was gathered and analyzed using the Statistical Package for Social Scientist (SPSS) and that of Microsoft Excel in a quantitative perspective.

## 5. RESULTS AND DISCUSSIONS

The structured questionnaires were divided into three categories. One for the workers of the DVLA, the Ghana Police Service, whilst drivers and vehicle owners taking the last category. The findings sought among other things; evasion of vehicle and driver license registration and related cost, the category of vehicles which are mostly stolen, reasons for reporting when your vehicle is stolen.

The research sought to know the most stolen category of vehicles or vehicles which are mostly reported as being stolen. This was sought for from the workers of DVLA and the MTTD officials of the Ghana Police Service. Private vehicles mysteriously dominated the first top three with a percentage of 56.8 representing a frequency of 25 of the total of 44 respondents. While motor bike which is noted to be used by many, came second with a frequency of 17 representing 38.6 percent, while passenger vehicle falling third with a percentage of 4.6 as shown in Table 1.

**Table 1: category of vehicles normally stolen**

| Type of vehicle   | Frequency | Percent      |
|-------------------|-----------|--------------|
| motor bike        | 17        | 38.6         |
| private vehicle   | 25        | 56.8         |
| passenger vehicle | 2         | 4.6          |
| <b>Total</b>      | <b>44</b> | <b>100.0</b> |

It became prudent to find out from the owners and drivers of vehicles to know their reasons for making reports when their vehicles are stolen. It was not strange to realize that about 88.9 percent of respondents representing a frequency of 80, report for the reason of recovering their vehicles. A frequency of 8, do report to meet insurance demands. Another frequency of 2, claim they report just for the purpose of assisting in the statistics gathering for the country. This could be seen in figure 1.

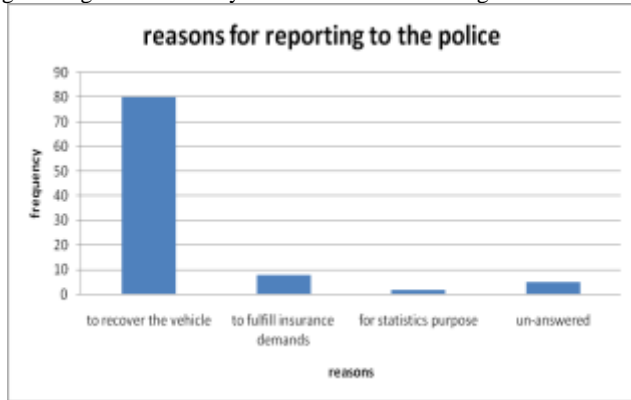


Figure 1: reasons for reporting a stolen vehicle

### 5.1 Evasion of Vehicle and Insurance Registration and Related Cost

Registration of all kinds relating to a vehicle is one of the ways government raises income to support the running of the state. But it is rather unfortunate that a number of people take to the business of evasion of licenses. From the survey conducted on the issue of the existence of fake license, out of 139 respondents, 120 attested to the fact that they have come across licenses which were fake as shown in Table 2. They gave several reasons for people embarking on such an activity. But here, strategically the views of DVLA staff were not sought for. Some of the reasons as seen in Figure 2 indicated 99 respondents out of a total of 133 were of the view that the fake license is cheaper to obtain, this represents 74.4 percent.

Table 2: Knowledge of fake license

| Response     | Frequency  | Percent      |
|--------------|------------|--------------|
| Yes          | 120        | 86.3         |
| No           | 19         | 13.7         |
| <b>Total</b> | <b>139</b> | <b>100.0</b> |

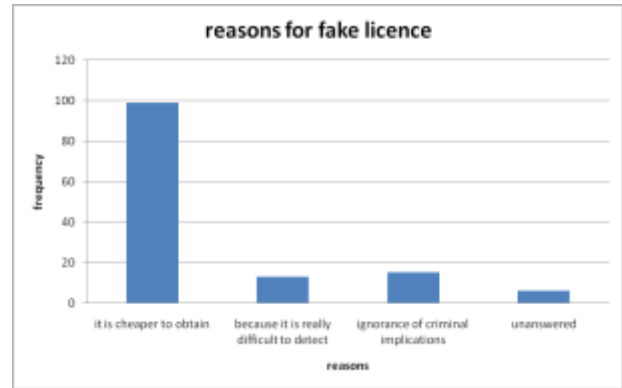


Figure 2: reasons for embarking on fake license registration

Inferring from the response from the Ghana Police Service, the most faked license in the country is the vehicle license, obtaining 52.6 percent of the 38 respondents, followed by the driver license with 31.6 percent and that of custom duty coming third with 15.8 percent as appeared in Table 3. There was no mention of insurance license, which presupposes that the insurance companies are getting it right.

Table 3: mostly faked license

| Type            | Frequency | Percent      |
|-----------------|-----------|--------------|
| vehicle license | 20        | 52.6         |
| driver license  | 12        | 31.6         |
| custom duty     | 6         | 15.8         |
| <b>Total</b>    | <b>38</b> | <b>100.0</b> |

There is always a cause for every action, thus we sought further from the drivers and vehicle owners their opinion for having to evade or fake license instead of getting the genuine one from the DVLA. It was revealing to note that cost was not the bone of contention but rather, the process to obtaining the license. 65.2 percent of 95 Respondents alluded to the fact that it is tedious to obtain a license. Whilst 31.6 percent were of the view that the registration is costly, this is illustrated in Figure 3.

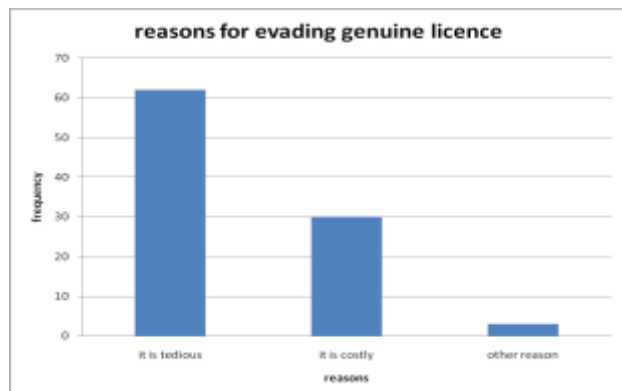


Figure 3: reasons for evading DVLA licence

It is thus necessary for putting in place a system of registration which will be fast and convenient for potential client’s patronage.

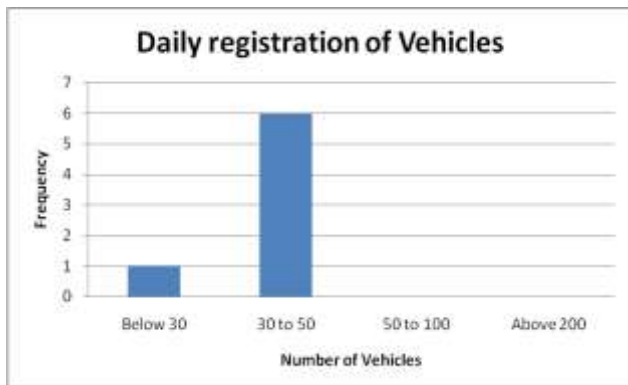
It could be noticed that the government earns a lot of income from the registration of vehicles. As can be seen from Table 4, a 58.9 percent representing 56 respondents indicated it cost them an amount more than GHS100 but less than GHS200 to register their vehicles. 25.3 percent of respondents indicated it cost them between GHS300 and GHS400 and a 5.3 percent said they spend more than GHS400.

**Table 4: Cost of Registering a Vehicle**

| Cost          | Frequency | Percent      |
|---------------|-----------|--------------|
| 100-200 GHS   | 56        | 58.9         |
| 200-300 GHS   | 24        | 25.3         |
| 300-400 GHS   | 10        | 10.5         |
| 400 and above | 5         | 5.3          |
| <b>Total</b>  | <b>95</b> | <b>100.0</b> |

It was necessary to know how much revenue is generated, thus we enquired from officials of the Driver and Vehicle Licensing Authority, how many vehicles they can register in a day. 6 respondents representing 85.7 percent said they register about 30 to 50 vehicle a day. While a respondent representing 14.3 percent indicated they register less than 30 vehicles a day as seen in Figure 4.

Inferring from Table 4 and Figure 4, it can be estimated that the daily revenue that could be generated will be an amount of GHS6,000(that is GHS150 by 40 vehicles), which is at least approximately GHS120,000 a month (considering averagely 20 working days in a month) and GHS1,440,000 for the year, only from the Tamale office of the Driver and Vehicle Licensing Authority.



**Figure 4: Number of vehicles registered daily**

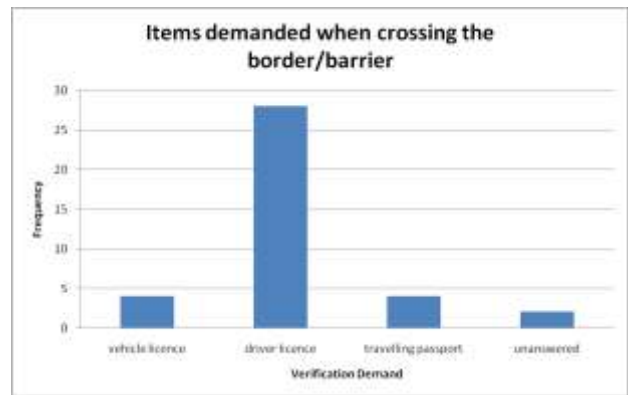
## 5.2 License Verification and Associated Technology

It was evident that verification is normally done in one way or the other by the Ghana Police Service on driver and vehicle license registration. The onus now lies on how it is being done. In responding to this as appeared in Table 5, 42 percent of a total of 38 respondents said they use the picture in the driver’s license card as a form of verification. All in all 63.1 percent of the respondents agreed they do normally verify, while a 31.6 percent are of the view that normally they do not verify, this I think they are not denouncing verification but that if there is no reported case, they will not undertake any verification.

**Table 5: verification of license for ownership by the Police**

| Ways of verification  | Frequency | Percent    |
|---|-----------|------------|
| We use the picture in the driver's license card             | 16        | 42         |
| We do cross check the vehicle license against DVLA register | 6         | 15.8       |
| We don't normally verify                                    | 12        | 31.6       |
| We verify if there is a reported case of a missing vehicle  | 2         | 5.3        |
| Unanswered  | 2         | 5.3        |
| <b>Total</b>  | <b>38</b> | <b>100</b> |

Observing critically how the police carried out their day to day verification for ownership of a vehicle, it was realized there still exist a big chance for thieves to do away with stolen vehicles especially getting their way to neighboring countries. This is because responses from the Ghana Police Service on what they require from people, who are driving across the border, indicated that driver’s license topped with a percentage of 73.7 out of a total of 38 respondents as seen in Figure 5.



**Figure 5: what police demand when crossing the border/barrier**

## 5.3 Summary of Findings

After the interpretation and analysis of the results, there were interesting and very revealing findings which are summarized here.

The Ghana Police noted they undertake license verification for ownership but this is normally carried out by the use of the driver's license.

There is the occurrence of Vehicle theft on a daily basis, snatching being the prominent means of perpetrating it. Private vehicles are the most stolen, and stolen vehicles are normally reported to the Police for the purpose of recovering such vehicles.

Also, respondents attest to the fact that there exist fake licenses in the country, with vehicle license being the most faked, blaming the tedious nature of getting a license on DVLA as the reason for the evasion.

## 6. RECOMMENDATIONS

The study makes the following recommendations as a result of the research findings gathered.

The government should empower the security services with a technology which can assist them to instantly track the licensing status of vehicles to boost the revenue generation capacity of the Driver and Vehicle Licensing Authority.

There should be a committed and zealous policy by the government to vigorously implement a technology to instantly verify the status of all kinds of licensing in the country.

There should be a national policy to educate the public on the economic and social implications of evading genuine licensing at the driver and vehicle licensing authority.

The government in collaboration with the Driver and Vehicle Licensing Authority (DVLA) should consider as a matter of urgency the automation of all the licensing processes at the DVLA to increase the patronage of clients.

## 7. REFERENCES

- [1] ABEO TIMOTHY APASIBA, J.B. HAYFRON-ACQUAH, J.K. PANFORD, MICHAEL ASANTE (2015). Enhancing the security of vehicles and the work of the DVLA using GSM/SMS technologies, International Journal of Science and Engineering Applications, Volume 4 Issue 5, 2015, ISSN 2319-7560 (Online)
- [2] AVRAHAM BRACHA (2008). *System and method for verifying Driver's insurance coverage*, U.S. Pat. US 2008/0027761 A1
- [3] AWUNI, MANASSEH AZURE (2011) *Can ACP Awuni change the face of MTU?* Available from: <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=202276> [Accessed 07/02/13]
- [4] BUSINESSGUIDEGHANA.COM (2012) *first capital plus pampers workers* Available from: <http://news.moneygh.com/pages/bankingandfinance/201206/50.php> [accessed 24/06/13]
- [5] CISSSE, L. (2006) *ECOWAS and the Daily Events: The Present Realities of the Integration Process within the Sub-Region* (GJA and FES: Accra), p.47.
- [6] EN.WIKIPEDIA.ORG (2012) *Interview*. Available from: <http://en.wikipedia.org/wiki/Interview> [Accessed 08/02/14]
- [7] ESERVICES.GOV.GH (2012) *online services*. Available from: <http://www.eservices.gov.gh/dvla/SitePages/DVLA-Home.aspx> [Accessed 30/05/13]
- [8] ESERVICES.GOV.GH (2013) *eservices portal of government of Ghana*. Available from: <http://www.eservices.gov.gh> [Accessed 26/05/13]
- [9] ESOKO.COM (n.d.) Available from: <http://www.esoko.com/about/index.php> [Accessed 21/06/13]
- [10] FIRSTCAPITALPLUS.NET (2011) Available from: <http://blog.firstcapitalplus.net/events/> [Accessed 24/06/13]
- [11] FLEETNEWS (2008) *DVLA launches e-licence checking service*. Available from: <http://www.fleetnews.co.uk/news/2008/12/11/dvla-launches-e-licence-checking-service/29530/> [Accessed 30/07/12]
- [12] Ghana News Agency (2007) *MTU gives statistics of fatal accidents*. Available from: <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=128741> [Accessed 07/02/13]
- [13] Ghana News Agency (2008) *DVLA registrations go online*. Available from: <http://www.modernghana.com/news/163669/1/dvla-registrations-go-online.html> [Accessed 21/06/13]
- [14] Ghana News Agency (2008) *road accidents cost Ghana Ghs 165,000 annually*. Available from: <http://www.modernghana.com/news/160253/1/road-accidents-cost-ghana-gh165000-annually.html> [Accessed 07/02/13]
- [15] Ghana Statistical Service (2009) *Victimization survey in Ghana*. Available from: [http://www.unodc.org/documents/data-and-analysis/dfa/Executive\\_Summary\\_GHANA\\_final.pdf](http://www.unodc.org/documents/data-and-analysis/dfa/Executive_Summary_GHANA_final.pdf) [Accessed 21/08/2010]
- [16] GHANABUSINESSNEWS.COM (2012) Available from: <http://www.ghanabusinessnews.com/2013/02/05/ashanti-dvla-raises-gh%2%A24-5m-revenue-in-2012/> [Accessed 07/02/13]
- [17] GHANAPOLICE.INFO (n.d.) Available from: [http://www.ghanapolice.info/broad\\_formation.htm](http://www.ghanapolice.info/broad_formation.htm) [Accessed 07/02/13]



- [18] Global Road Safety Partnership (2012) Available from: [www.grsproadsafety.org](http://www.grsproadsafety.org) [Accessed 06/01/13]
- [19] GNA (2012) *EC to Launch “BVR SMS Checking” on 3<sup>rd</sup> September*, Available from: [http://www.africanelections.org/ghana/news\\_detail.php?nws=6560&t=EC%20to%20launch%20%E2%80%9CBVR%20SMS%20Checking%E2%80%9D%20on%203rd%20September](http://www.africanelections.org/ghana/news_detail.php?nws=6560&t=EC%20to%20launch%20%E2%80%9CBVR%20SMS%20Checking%E2%80%9D%20on%203rd%20September) [Accessed 21/06/13]
- [20] JIWA ABDULLAH (2011) the design of mobile control car security system, *International Journal of Engineering and Technology* IACSIT, Vol.3, No.3, p.5
- [21] JOSHUA PEPRAH (2008). Cell Phone Penetration to Hit 50%, Available from: <http://www.ghanaweb.com/GhanaHomePage/economy/artikel.php?ID=144001>
- [22] KHONDKER SHAJADUL HASAN, MASHIUR RAHMAN, ABUL L. HAQUE, M ABDUR RAHMAN, TANZIL RAHMAN, M MAHBUBUR RASHEED (2009) Cost Effective GPS-GPRS Based Object Tracking System, *Proceedings of the International MultiConference of Engineers and Computer Scientists* Vol I, p.6 IMECS 2009, March 18 - 20, 2009, Hong Kong
- [23] MAHESH KUMAR SHAKYA (2012) Catching Of Stolen Vehicles With Unique Identification Code Using Embedded Systems, *International Journal of Electronics Communication and Computer Technology* (IJECCCT) Volume 2 Issue 6, p.5
- [24] MILAN ZLOJUTRO (2010) *Vehicle Tracking and Monitoring System* U.S. pat. US 7,701,363 B1, p.2
- [25] OSAC (2013) *Ghana 2013 Crime and Safety Report*. Available from: <https://www.osac.gov/pages/ContentReportDetails.aspx?cid=13768> [Accessed 15/3/2013]
- [26] SHIHAB A. HAMEED, SHAIMA ABDULLA, AISHA HASSAN, OTHMAN KHALIFA, JAMAL I. DAOUD (2012) Effective Car Monitoring And Tracking Model, *Australian Journal of Basic and Applied Sciences*, 6(1): pp. 1-8, 2012

# Novel Approach for Card Payment

Amrapali.Patil  
MIT College of  
Engineering,Pune  
Pune,India

Piyush.Potdar  
MIT College of  
Engineering,Pune  
Pune,India

Swapnaja.Lomte  
MIT College of  
Engineering,Pune  
Pune,India

**Abstract:** The ultimate across the board user authentication approach in use today is evidently the password-based authentication. When we carry out a credit card transaction through the EDC (Electronic Data Capture) machine in the public, the user's PIN number becomes very much vulnerable to the direct observation by nearby adversaries in huddled places, promoted by vision enhancing and/or recording appliances. Devising a secure PIN entry method during the credit card transaction in such a situation is a strenuous task. Currently, there is no pragmatic solution being implemented for this problem. This paper starts with the investigation of the current status about the direct experiential attacks. Our analysis about these attacks terminates that no practical available solution at present for these direct observational attacks. This paper introduces a model which attempts to make the PIN number entry secure during credit card transactions in public places. Our model aims to use the user's mobile phone for PIN number entry rather than the merchant's user machine. The best tract about the proposed model is that the PIN number does not get revealed to any of the direct observational attacks, be it direct human observation or observation by a video camera.

**Keywords:** overlooking, shoulder-surfing, chip and pin, bank server, user's PIN.

## 1. INTRODUCTION

The chip and Pin method is carried out as follows. The Merchant or the public place where the user is paying his/her bill presents the user with the EDC machine. The user swipes his/her credit card through the EDC machine. The credit card specifics get recorded in the EDC machine. Then the user will enter his/her PIN number through the keypad on the EDC machine. The credit card details captured by the machine are matched with the PIN number. If matched, the bank will allow the merchant to carry out the transaction. For committing a credit card fraud, two things are very Crucial, one is the credit card number and the other is the PIN number. Obtaining the credit card number has become very easy these days. Credit card number can easily be obtained from the receipt produced during any credit card transaction. Or it can be recorded by a fraud person by installing a card reader in the EDC machine. So, if the credit card number is so effortless to obtain, it all comes down to the PIN number. Hence, securing the PIN number becomes very critical. Even partial information about the PIN number if leaked, can prove to be harmful as users tend to use identical passwords on multiple systems. Or, as the information is partially known, guessing the remaining part may become easier.

The problem in the chip and pin method arises during the credit card transaction when the user is entering the PIN number in a public or a crowded place such as, restaurant or a shopping mall. In this situation, the user is obliged to enter his/her PIN number in front of all those around him in the surrounding. Hence, the PIN number becomes susceptible to the direct observational attacks by humans. Or there can be a video camera or others such device which can easily capture the user entering PIN number through the EDC machine. If a user performs the credit card transaction regularly at some place, for example at a cafeteria every day, the employee by observation can learn about the users PIN number eventually. Or it may get recorded in the video Cameras at the public places. So, the PIN number is being exposed in a relatively non-technical aspect. Such kind of attacks are called overlooking or shoulder-surfing attacks. The revealing of passwords in this manner has become a security concern today due to the many cases of frauds being committed in this manner.

After the introduction of Chip-and-Pin method magnetic stripes cards were replaced with EMV cards, Chip-and-Pin method drastically reduced fraud rates in US and Europe to a certain level. This method may have reduced frauds in face-to-face transaction but we think this is not sufficient and with the implementation of our algorithm and setup this can be further reduced. Fraud losses associated with face-to-face transaction in UK itself were 198 million pounds in financial year 2008 which is pretty huge amount while the total fraud losses in UK in same year were skyrocketing high i.e. 600 million pound. World is now adapted to Chip-and-Pin method but this technique is not totally safe, as per the federal bank records in 2008 in USA fraud losses to all parties on card transaction per dollar volume were 0.13 percent or 13 basis point. This figures tells us that a lot of work can be done to reduce this losses, so we proposed a method which is capable of this. As of now, no solution is being implemented. There are some methods like BW (Black White) scheme, TictocPIN proposed for the overlooking or shoulder attacks but none of them have yet been implemented for practical use due to their respective drawbacks. So, at present all the user can do to secure his/her PIN number from being revealed is that, cover

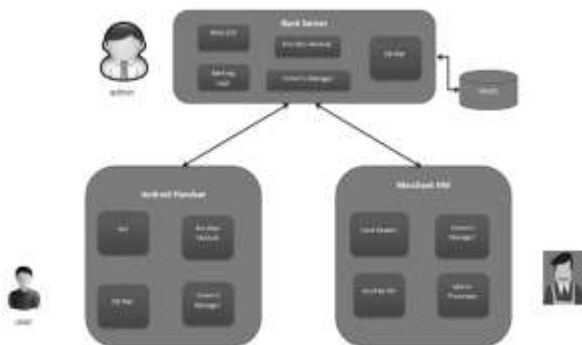


Figure 1. Schematic diagram of proposed system Model.

it while PIN number entry through the EDC machine with his/her hand or Also, review the statements after the transaction. For such attacks, we propose a model which defies overlooking or shoulder-surfing attacks. Our model uses the Users cell phone for PIN number entry instead of the user's Cell phone.

## 2. LITERATURE SURVEY

The solutions for making the PIN number entry during the credit card transactions are many. However, none of these implementations have brought into practical use yet. Because some have drawbacks in security whereas some others in the usability. This indicates the difficulty in securing the PIN number during offline credit card transaction.

Previously, methods have been suggested which use graphical passwords [3] [4], textual passwords [5] [6] as well a method which uses the combination of both [6]. In the method proposed by H. J. Asghar et al. [7], the user has to perform a convex hull of secret icons mentally in a set of graphical icons and then click randomly within this hull. In PAS i.e. predicate based authentication proposed by X. Baiet.al [5], predicates are used for authentication. The method described in [8] use haptic and audio signals as input for PIN number entry as these type of inputs tend to be resilient against camera attacks. The method ColorPIN [9] uses coloured letter to indicate the digits in the PIN number. But this method is not resilient against camera attacks. In one another method, pressure grids [10] are used to indicate the digit in the PIN number. However, this method has usability issues, especially in the time and efforts of the user to enter the PIN. In the method proposed by Roth et. Al [1] referred to as the BW method in [2] the user uses a simple keypad layout with half of the keys coloured white and the other half, black for entering the PIN number. The user gives input by pressing on the colours of his digit in the PIN. So the user does not give the actual digits but indicates them by colour. The entry of the PIN takes 16 rounds. It is a simple method but has practical usability issues like round redundancy, unbalanced frequencies and little protection against recording attacks [2] This method fails if sophisticated strategies and training adopted by adversaries. An improvement over this BW scheme is proposed by Kwon and Jin Hong [2] called as TictocPIN. This method assigns colours to the numeric keys. Four and three subsets of numbers are created. Unlike BW, here fixed partitions of digit space are to be used in all the sessions. It aims to receive each PIN digit through multi-round challenge round procedure. The user is informed through short vibration signals as to which of Multiple displayed challenges is to be taken as valid. This method provides sufficient security against even camera based Recording attacks. The drawback is that the time required may be considered as uncomfortable for users use.

## 3. OVERVIEW OF MODEL

The model aims to take the PIN entry through his mobile phone rather than the conventional method of using the same EDC machine. One of the purposes of doing this is to provide more closure to the user while entering PIN. The user swipes his/her credit card through the EDC machine. Then the

merchant notifies the bank about this swipe. The bank sends notification on the user's mobile phone. The user enters his PIN through his mobile phone instead of the EDC machine. In one kind of attack, the adversary may install a card reader in the EDC machine and when the user swipes his/her card or enters the PIN number, the user's details may get recorded in the card reader. The method of entering the PIN number through the user's mobile phone deflects this possibility of such an attack. It can be called as Man-in-the-middle attack.

When at a public or cramped place the user makes a prime of paying his/her bill through the credit card the user is conferred with the merchants EDC machine through which the user swipes his/her credit card. This type of transaction is often called as the offline credit card transaction. On swiping, the credit card details are captured by the EDC machine. These specifics will be sent by the merchant to the bank through the Internet. The bank gets notified about the card swipe. The bank will immediately send a notification on the user's cell phone. The user's cell phone has an application which takes the PIN entry input. The application on the user's cell phone is the most crucial part of this model. The application aims to not to reveal the users PIN in the public place which means the user will not enter the PIN at all in the entire transaction. So, the PIN entry will be made in the application by a manner which will satisfy such condition. There are many different Methods of achieving this.

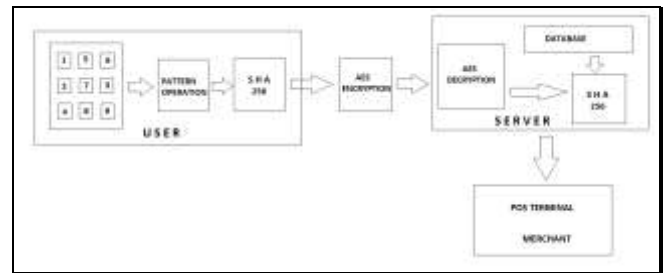


Figure 2. Detailed System Architecture of proposed Model.

The user will have the flexibility to choose any method which he/she finds suitable. Some of the methods are as follows. One way is through pattern entry like the one that is available for the phone lock. So when the user gets a notification on the mobile phone, he will enter the pattern and the PIN number stored in the mobile phone in the encrypted form will be sent to the bank. The advantage here is that the pattern is only known to the user. Also, even if it faces shoulder-surfing or overlooking attack or gets recorded by a video camera, will be of no use because knowing the pattern is basically useless. Unless the PIN number is known, the efforts of the adversary are futile. Another way of pattern entry could be to have a password entry which again will have the same advantage as the previous method. Or one can use other methods like reverse pattern or half reverse pattern. If the user uses reverse pattern, only the user knows that the PIN number to be entered has to be in reverse. For example, if the PIN number is 0793, the user will enter 3970. So the adversaries will consider the users PIN number to be the 3970

Whereas the actual PIN being reverse of it. The entered PIN after pattern analysis will be sent in an encrypted manner to the bank. The half pattern will work in the similar manner. Here, the user will have to enter 7039 i.e. half of the digits in the reverse order. One can use many such type of patterns which after pattern analysis by the application will be

converted back to the actual PIN. Thus, the actual PIN does not get revealed during the Pin number entry. Or a simpler method would be to use a yes/no pattern i.e. when the bank notifies the user about the card swipe, the user will simply press on either yes/no provided by the application. On entering yes, the PIN stored in the phone in the encrypted form will be sent to the bank. One can also use the TicTocPIN scheme proposed by Kwon and Jin Hong [2] on the phone application as it is by far the most viable method proposed. The user has a choice to use the PIN entry method of his/her choice, whichever he/she finds appropriate and secure.

### 3.1 Detailed Working of Model

The main purpose of this application is to keep the PIN hidden from observers and cameras. In some of the methods above, the PIN number needs to be hoarded in the phone in an encrypted manner. For security, this PIN number will be stored in the phone using SHA (secure hash algorithm) coding. SHA 256 is a cryptographic hashing function. It allows an almost unique, fixed size 256 bit (36 byte) .It is a one way function. Once operated, it cannot be decrypted back. This makes it suitable for password validation. So even if the phone comes into the hands of the adversary, the SHA coded PIN will be of no use .And if the PIN number is not stored in the phone i.e. it is entered every time during the transaction using some pattern, then the entered PIN after pattern analysis will be converted back to the actual PIN which in turn will be operated with SHA. After the SHA coding of the PIN number, It needs to be delegated to the bank, through the network. To make the PIN more secure, the SHA coded is AES encrypted. Or any such similar encryption algorithm can be used. After AES encryption, it will be directed through the network to the bank. At the bank end, the authentication will be carried out in the following manner. The incoming PIN number from the user through the network will first be decrypted back to the SHA coded PIN. As the bank database will have users PIN number stored, it will get the PIN of that particular user. The PIN will undergo SHA coding as the incoming PIN from the user is SHA coded. The incoming SHA coded PIN is matched with that from the database. Depending on the match results, the bank allows the merchant to carry out the supplementary

### 4. CONCLUSION

Through the proposed model, the vital objective of keeping the PIN confidential from the observer or video cameras is accomplished. PIN does not get revealed at any stage of the transaction. Thus, the model is resilient to overlooking or shoulder-surfing attacks. As the application gives the user the flexibility to choose the PIN entry method according to his/her usability and security, this makes it strenuous for the adversaries to obtain the PIN number by sophisticated approach.

Also, as the users mobile phone provides more closure to the user than the EDC machine, even observation along with the camera recording attacks become difficult for the adversaries. The bank notifying the user after the card swipe by the user serves another purpose. When the card is swiped, the bank is notified about the card swipe. The bank in turn notifies the user about the card swipe and asks for the PIN entry .In case of the card gets stolen and is used by the adversary for swiping at some place, the user will immediately be notified about the card swipe. This serves to authenticate the card

owner. In case of the stolen card, the user due to the bank notification will easily come to realize about his/her stolen card if not known until then. The user can then take suitable action as soon as possible to avoid any damage. The model also delivers to make attacks other than the overlooking, shoulder-surfing or recording attacks resilient. One such attack is man-in-middle attack. If a card reader is installed in the EDC machine, all the credit card specifics entered through the EDC machine will be captured by the adversaries. In our proposed model, as the EDC machine is not used for the PIN number entry, capturing of the data by the third party is averted. The PIN number remains secure making the man-in-middle attack futile.

The proposed model has Internet dependency. In case there is no Internet available, there will constantly be an option for the user to enter the PIN number in the popular manner as our model does not change any aspects of the conventional method of the offline credit card transaction.

### 5. REFERENCES

- [1] V. Roth , K. Richter, and R. Freidinger, “A PIN-entry method resilient against shoulder surfing”, in Proc. 11th ACM Conf. Comput. Commun. Secur. (CCS), 2004, pp. 236–245.
- [2] Taekyoung Kwon ; Grad. Sch. of Inf., Yonsei Univ., Seoul, South Korea ; Jin Hong “Analysis and Improvement of a PIN-Entry Method Resilient to Shoulder-Surfing and Recording Attacks” in Information Forensics and Security, IEEE Transactions on (Volume:10 , Issue: 2 ) Feb. 2015 pp. 278 – 292.
- [3] P. Dunphy, A. P. Heiner, and N. Asokan, “A closer look at recognition based graphical passwords on mobile devices” in Proc. 6th Symp. Usable Privacy Secur. , 2010, pp. 1–12.
- [4] Q. Yan, J. Han, Y. Li, and R. H. Deng, “On limitations of designing leakage-resilient password systems: Attacks, principals and usability” in Proc. 19th Symp. Internet Soc. Netw. Distrib. Syst. Secur. (NDSS), Feb. 2012. Brown, L. D., Hua, H., and Gao C. 2003. A widget framework for augmented interaction in SCAPE.
- [5] X. Bai, W. Gu, S. Chellappan, X. Wang, D. Xuan, and B. Ma, “PAS: Predicate-based authentication services against powerful passive adversaries,” in Proc. IEEE Annu. Comput. Secur. Appl. Conf., Dec. 2008, pp. 433–442.
- [6] H. Zhao and X. Li, “S3PAS: A scalable shoulder-surfing resistant textual-graphical password authentication scheme,” in Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl. Workshops, vol. 2. May 2007, pp. 467–472.
- [7] H. J. Asghar, S. Li, J. Pieprzyk, and H. Wang, “Cryptanalysis of the convex hull click human identification protocol,” in Proc. 13th Int. Conf. Inf. Secur., 2011, pp. 24–30.
- [8] A. Bianchi, I. Oakley, and D. S. Kwon, “Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry,” *Interact. Comput.* , vol. 24, no. 5, pp. 409–422, Sep. 2012.

[9] A. De Luca, K. Hertzschuch, and H. Hussmann, “ColorPIN—Securing PIN entry through indirect input,” in Proc. ACM CHI Conf. Human Factors Comput. Syst., 2010, pp. 1103–1106.

[10] D. Kim et al., “Multi-touch authentication on tabletops,” in Proc. ACM SIGCHI Conf. Human Factors Comput. Syst. (CHI), 2010, pp. 1093–1102.



# An Inventory Model with Variable Demand Rate for Deteriorating Items under Permissible Delay in Payments

Ajay Kumar Agrawal  
Motilal Nehru College  
University Of Delhi New Delhi,  
India

Kuldeep  
Department Of Mathematics  
Lady Shri Ram College For Women  
University Of Delhi  
New Delhi, India

---

**Abstract:** A continuous production control inventory model for deteriorating items with variable demand rate is developed. Demand rate is the linear function of time. In this paper we have done all work in the environment of permissible delay of payments. A number of structural properties of the inventory system are studied analytically. We have discussed the minimum total system cost under the condition of permissible delay is relaxed to that at the end of the credit period, the retailer will make a partial payment on total purchasing cost to the supplier and pay off the remaining balance by loan from the bank. Numerical examples are taken to illustrate the procedure of finding the optimal total inventory cost and production cycle time. Sensitivity analysis is carried out to demonstrate the effects of changing parameter values on the optimal solution of the system.

**Keywords:** Inventory, EPQ Model, Deteriorating item, variable demand rates, permissible delay in payment.

---

## 1. INTRODUCTION

The classical Economic Production Quantity (EPQ) model is widely used principally because it is so simple to use and apply. This model considers the ideal case that the value of inventory items are unaffected by time and replenishment is done instantaneously. In real life cases, however, the ideal case is not quite applicable. Inventories are often replenished periodically at certain production rate which is seldom infinite. Even for purchased items, when supply arrives at the warehouse, it may take days for receiving department to completely transfer the supply into storage room. Goods deteriorate and their value reduces with time. Electronic products may become obsolete as technology changes: Fashion tends to depreciate the value of clothing over time; batteries die out as they age. The effect of time is even more critical for perishable goods such as food stuff and cigarettes. The effect of these two situations is that the classical inventory model has to be readjusted.

The traditional Economic Order Quantity (EOQ) model assumes that the retailer must be paid for the items as soon as the items are received. However, in practice the supplier will offer the retailer a delay period, that is the trade credit period, in paying for the amount of purchasing cost. Before the end of the trade credit period the retailer can sell the goods and accumulate revenue and earn interest. A higher interest is charged if the payment is not settled by the end of trade credit period. Therefore, it is clear that the retailer will delay the payment up to the last moment of the permissible period allowed by the supplier. In the real world, the supplier often makes use of this policy to stimulate his/her customer's demand. Recently several papers

have appeared in the literature that treat inventory problems with varying conditions under the consideration of permissible delay in payments. Some of the prominent papers are discussed below.

Goyal [1] established a single-item inventory model under permissible delay in payments. Chung [2] developed an alternative approach to determine the economic order quantity under condition of permissible delay in payment. Aggarwal and Jaggi [3] considered the inventory model with exponential deterioration rate under the condition of permissible delay in payments. Jamal et al [4] extended Aggarwal and Jaggi's model to shortages. There were several interesting and relevant papers related to the delay of payments such as Chu et al [5], Chung [6], Hwang and Shinn [7], Sarker et al [8], Shah [9], Shinn [10], Khouja and Mehrez [11] and their references. However, these studies were developed under the assumption that the items obtained from an outside supplier and the entire lot size is delivered at the same time, when an item can be produced inhouse, the replenishment rate is also the production rate and is finite. Hence we amend Goyal's model by considering the replenishment rate is finite, the difference between purchasing price and selling cost and taking into consideration the effect of time or decay.

Specifically, the restrictive assumption of a permissible delay is related to that at the end of the credit period, the retailer will make a partial payment on total purchasing cost and pay the remaining balance by loan from the bank.

1. In reality, the demand may vary with time. Time-varying demand patterns are commonly used to reflect sales in different phases of a product life cycle in market. For example, the demand for inventory items

increases over time in the growth phase and decrease in the decline phase. Donaldson [15] initially developed an inventory model with a linear trend in demand. After that many researcher's works in this environment of Goal and Aggrawal [17], Ritchie [21] Deb and Chaundhari [14]. Dave and Palel [13], Chung and Ting [12], Kishan and Mishra [19] Giri et al [16], Hwang [18], Pal and Mondal [20], have be devoted to incorporating a time-varying demand rate into their models for deteriorating items. Consequently, the main purpose of this paper is to find an optimal cycle time which minimize the total system cost. Numerical examples are also presented to illustrate the result of the proposed model.

**Notation**

|                    |   |
|--------------------|---|
| D                  | Annual demand rate  |
| P                  | Annual production rate  |
| c                  | Unit purchasing price per item  |
| s                  | Unit selling price per item of good quality                               |
| h                  | Unit stock holding cost per item per year excluding interest changes      |
| Ie                 | Interest which can be earned per \$ per year                              |
| Ip                 | Interest changes per \$ investment in inventory per year                  |
| I <sub>1</sub> (t) | The inventory level that changes with time t during production period     |
| I <sub>2</sub> (t) | The inventory level that changes with time t during non-production period |
| M                  | The trade credit period   |
| T                  | The cycle time  |
| T*                 | The optimal cycle time  |
| θ                  | Deterioration rate of finished item                                       |
| A                  | Setup cost per year   |

**Assumption:**

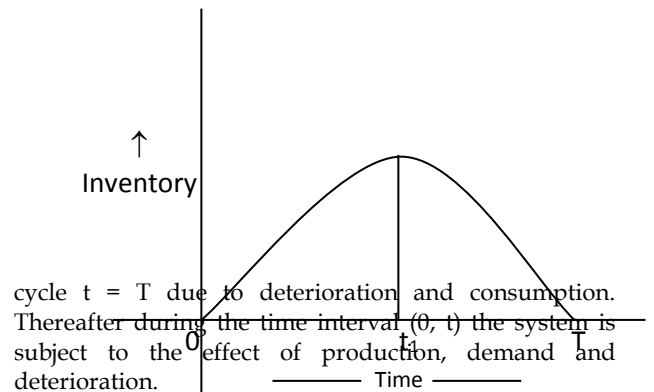
1. Production rate P is known and constant
2. Demand rate  $D(t) = bt$ ,  $b > 0$  and  $P > D(t)$  for any t
3. Shortages are not allowed
4. Time period is infinite
5. Lead time is zero
6. The constant fraction  $\theta$  of on hand inventory gets deteriorated per time unit
7.  $s \geq c$ ,  $Ip \geq Ie$

**3. MATHEMATICAL MODEL AND SOLUTION**

A constant production rate starts at  $t = 0$  and continuous up to  $t = t_1$  where inventory level reaches the maximum level. Production then stops at  $t = t_1$  and the inventory gradually depletes to zero at the end of the production

**2. ASSUMPTIONS AND NOTATIONS**

All The proposed inventory model is developed under the following assumptions and notations.



$$\frac{dI_1(t)}{dt} = P - D(t) - \theta I_1(t) \quad 0 \leq t \leq t_1$$

Or

$$\frac{dI_1(t)}{dt} = P - bt - \theta I_1(t) \quad 0 \leq t \leq t_1 \quad \dots(1)$$

With initial condition  $I_1(0) = 0$

On the other hand, in the time interval  $(t, T)$ , the system is affected by the combined effect of demand and deterioration. Hence, the change in the inventory level is governed by the following differential equation.

$$\frac{dI_2(t)}{dt} = -D(t) - \theta I_2(t) \quad t_1 \leq t \leq T$$

Or

$$\frac{dI_2(t)}{dt} = -bt - \theta I_2(t) \quad t_1 \leq t \leq T \quad \dots (2)$$

With ending condition  $I_2(T) = 0$

The solution of the differential equation (1) and (2) are respectively represented by

$$I_1(t) = \left( \frac{P}{\theta} + \frac{b}{\theta^2} \right) \left( 1 - e^{-\theta t} \right) - \frac{bt}{\theta}, \quad 0 \leq t \leq t_1 \quad \dots(3)$$

And

$$I_2(t) = \frac{b}{\theta^2} \left[ (\theta T - 1) e^{\theta(T-t)} - \theta t + 1 \right] \quad \dots(4)$$

In addition, from the boundary condition  $I_1(t_1) = I_2(t_1)$  we can derive the following equation.

$$\left(\frac{P}{Q} + \frac{b}{Q^2}\right)e^{-\theta t_1} = \frac{P}{Q} - \frac{b}{Q^2}(QT-1)e^{\theta(T-t_1)} \quad \dots(5)$$

and

$$t_1 = \frac{1}{\theta} \log \left[ \frac{\left(\frac{P}{Q} + \frac{b}{Q^2}\right) + \left(\frac{bT}{Q} - \frac{b}{Q}\right)e^{\theta T}}{P} \right] \quad \dots(6)$$

For the moment, the individual costs are now evaluated before they are grouped together

1. Annual setup cost =  $\frac{A}{T}$

2. Annual holding cost (excluding interest charges)

$$\begin{aligned} &= \frac{h}{T} \left[ \int_0^{t_1} I_1(t) dt + \int_{t_1}^T I_2(t) dt \right] \\ &= \frac{h}{\theta^3 T} \left[ (P\theta + b)e^{-\theta t_1} + b(\theta T - 1)e^{\theta(T-t_1)} + \theta^2(Pt_1 - bT^2) - P\theta \right] \end{aligned}$$

#### 4. INTEREST

Case I: when  $T \leq M$

(a) **Interest earned:** In this case, the customer sells D.T unit in total by the end of the replenishment cycle time T, and has cDT to pay the supplier in fall by the end of the credit period M, consequently, there is no interest payable. However, the interest earned per year is

$$\begin{aligned} &= sI_e \left[ \int_0^T D.t dt + (M-T) \int_0^T D dt \right] / T \\ &= \frac{1}{6} sbI_e (3M - T)T \end{aligned}$$

(b) **Interest charges**

In this case no interest charges are paid for the items kept in stock.

Case II: when  $M \leq T$

(a) **Interest earned:**

During the permissible delay period, the buyer sells products and deposits the revenue into an account that earns Ie per dollar per year

Therefore, the interest earned per year is

$$\begin{aligned} &= sI_e \int_0^m D.t \frac{dt}{T} \\ &= \frac{sI_e b M^3}{3T} \end{aligned}$$

(b) **Interest charged**

The buyer sells D.M in total by the end of the permissible delay M and has cDM to pay supplier. The item in stock are charged at interest rate Ip by the

supplier starting at the time M. Thereafter the buyer gradually reduces the amount of financed loan from supplier due to constant sales and revenue received. As a result the interest payable per year

$$\begin{aligned} &= \frac{cIp}{T} \int_M^T I_2(t) dt \\ &= \frac{bcIp}{T\theta^2} \left[ \left(T - \frac{1}{\theta}\right)e^{\theta(T-M)} - \frac{1}{2}\theta(T^2 - M^2) - M + \frac{1}{\theta} \right] \end{aligned}$$

#### 5. ANNUAL COST DUE TO DETERIORATED UNITS

$$\begin{aligned} &= \frac{c}{T} \left[ Pt_1 - \int_0^{t_1} D(t) dt - \int_{t_1}^T D(t) dt \right] \\ &= c \left( \frac{Pt_1}{T} - \frac{1}{2} bT \right) \end{aligned}$$

Therefore, the total variable cost function per unit time TVC(T) is

$$TVC(T) = \begin{cases} TVC_1(T) & \text{if } M \leq T \\ TVC_2(T) & \text{if } 0 \leq T \leq M \end{cases} \quad \dots(7)$$

And

TVC(T) = Setup cost + Stock holding cost + interest payable - interest earned + Annual cost due to deteriorated units

Then,

$$\begin{aligned} TVC_1(T) &= \frac{A}{T} + \frac{h}{\theta^3 T} \left[ (P\theta + b)e^{-\theta t_1} + b(\theta T - 1)e^{\theta(T-t_1)} + \theta^2(pt_1 - bT^2) - P\theta \right] \quad \dots(8) \\ &+ \frac{bcIp}{\theta^2 T} \left[ \left(T - \frac{1}{\theta}\right)e^{\theta(T-M)} - \frac{1}{2}\theta(T^2 - M^2) - M + \frac{1}{\theta} \right] - \frac{sI_e b M^3}{3T} + c \left( \frac{Pt_1}{T} - \frac{1}{2} bT \right) \end{aligned}$$

$$\begin{aligned} TVC_2(T) &= \frac{A}{T} + \frac{h}{\theta^3 T} \left[ (P\theta + b)e^{-\theta t_1} + b(\theta T - 1)e^{\theta(T-t_1)} + \theta^2(pt_1 - bT^2) - P\theta \right] - \frac{1}{6} sbI_e (3M - T)T \quad \dots(9) \\ &+ c \left( \frac{Pt_1}{T} - \frac{1}{2} bT \right) \end{aligned}$$

Since  $I_1(t_1) = I_2(t_1)$  and approximate value of  $e^{\theta(T-M)} = 1 + \theta(T-M) + \frac{\theta^2(T-M)^2}{2}$  which implies equation

(8) and (9) can be rearranged as following:

$$TVC_1(T) = \frac{A}{T} + \frac{h}{\theta T} (Pt_1 - bT^2) + \frac{1}{2} bcIp(T-M)^2 - \frac{sI_e b M^3}{3T} + c \left( \frac{Pt_1}{T} - \frac{1}{2} bT \right) \quad \dots(10)$$

And

$$TVC_2(T) = \frac{A}{T} + \frac{h}{\theta T} (Pt_1 - bT^2) - \frac{1}{6} sbI_e (3M - T)T + c \left( \frac{Pt_1}{T} - \frac{1}{2} bT \right) \quad \dots(11)$$

The objective in this paper is to find an optimal cycle time to minimize the total variable cost per unit time.

For this, the optimal cycle time  $T_1^*$ , obtained by setting

the derivative of equations (10) with respect to T equal to zero is the root of the following equation

i.e,

$$\frac{dTVC_1(T)}{dT} = -A + \left(\frac{C\theta + h}{\theta}\right)P \left(T \frac{dt_1}{dT} - t_1\right) - b \left(\frac{h}{\theta} + \frac{c}{2}\right)T^2 + bcIp(T-M)T^2 + \frac{slebM^3}{3} = 0 \dots(12)$$

And

$$\frac{d^2TVC_1(T)}{dT^2} = \left(\frac{c\theta + h}{\theta}\right)PT \frac{d^2t_1}{dT^2} - b \left(\frac{2h}{\theta} + c\right)T + bcIp(3T^2 - 2TM) \dots(13)$$

Where

$$\frac{dt_1}{dT} = \frac{Pb\theta T e^{\theta T}}{(P\theta + b) + b(\theta T - 1)e^{\theta T}}$$

And

$$\frac{d^2t_1}{dT^2} = \frac{Pb\theta \left[ (Pb\theta T + PQ + \theta^2 T + b)e^{\theta T} - be^{2\theta T} \right]}{\left[ (P\theta + b) + b(\theta T - 1)e^{\theta T} \right]^2}$$

In this same way, the optimal cycle time  $T_2^*$ , obtained by setting the derivative of equation (11) with respect to T equal to zero is the root of the following equation

$$\frac{dTVC_2(T)}{dT} = -A + \left(\frac{c\theta + h}{\theta}\right)P \left(T \frac{dt_1}{dT} - t_1\right) - b \left(\frac{h}{\theta} + \frac{c}{2}\right)T^2 - \frac{1}{6}sble(3M - 2T)T^2 \dots(14)$$

And

$$\frac{d^2TVC_2(T)}{dT^2} = \left(\frac{c\theta + h}{\theta}\right)PT \frac{d^2t_1}{dT^2} - b \left(\frac{2h}{\theta} + c\right)T - bsle(M - T)T \dots(15)$$

In order to find the optimal value of  $T_1$  and  $T_2$  i.e.,  $T_1^*$  and  $T_2^*$  so that the total variable cost per unit time is minimized i.e., evaluate that values of  $T_1$  and  $T_2$  from equation (12) and (14) so that

$$\frac{d^2TVC_1(T)}{dT^2} > 0 \text{ and } \frac{d^2TVC_2(T)}{dT^2} > 0$$

## 6. SOLUTION PROCEDURE

The optimal solution of the system can be obtained from the following algorithm

### Algorithm

Step 1: Using all the system parameters in equations (12) and (14)

Step 2: Find out all possible values of T

Step 3: Choose the optimal value of T so that

$$\frac{d^2TVC_1(T)}{dT^2} > 0 \text{ and } \frac{d^2TVC_2(T)}{dT^2} > 0$$

Step 4: Using the optimal values of T in (10), (11) and (6) and find the total variable cost minimum and production time  $t_1$ .

## 7. NUMERICAL EXAMPLES

To illustrate the results let us apply the proposed method to solve the following numerical examples. The following parameters  $h = \$ 5/\text{unit}$ ,  $C = \$ 5 / \text{units}$ ,  $S = \$ 6 \text{ units}$ ,  $b = 1000$ ,  $I_e = \$ 0.1 / \$ / \text{year}$ ,  $I_p = \$ 0.15 / \$ / \text{year}$  and  $Q = 0.3$  are used in appropriate units. If  $A = \$ 50 / \text{order}$ ,  $P = 5000 \text{ unit/unit time}$  Then for  $T = 3.42$ ,  $TVC_1 = \$ 1250$ .

## 8. CONCLUSION

This study presents a production inventory model for deteriorating items under permissible delay in payments with variable demand rate and small deterioration rate. In this work we found an optimal cycle time which give minimum total variable cost of the system corresponding to different value of parameters. This work done with linear demand rate. Numerical examples reveal that our optimization procedure is very accurate and rapid. From sensitive analysis we got sufficient variation in the value of parameters.

Briefly, results in this paper is better than the results of that works in which taken constant demand rate. Finally, a future study will incorporate more realistic assumption in the proposed model, such as variable deterioration rate, stochastic nature of demand and production rate which depends on both on-hand inventory and demand.

## 7. REFERENCES

1. S.K. Goyal, Economic order quantity under conditions of permissible delay in Payments. j. operational Res. Soc 36 (1985) 335-338.
2. K.J Chung, A theorem on the determination of economic order quantity under conditions of permissible delay in payments. Comput. Oper Res 25(1998) 49-52.
3. Aggarwal , S.P. and C.K. Jaggi, 1995. Ordering policies of deteriorating items under permissible delay in payment. j. operational research SOC 46, 658-662.
4. A.M.M. Jamal B.R. Sarker, S. Wang. Optimal Payment time for a retailer under permitted delay of payment by the wholesaler. Int. J. Prob. Econ 66(2000) 59-66.
5. P.Chu, K.J Chung, S.P. Lan, Economic order quantity of deteriorating items under permissible delay in payments. Comput. Oper . Res. 25(1998), 817-824.
6. K.J Chung, The inventory replacement policy for deteriorating items under permissible delay in payments. Oper. Res. 37(2000) 267-281.
7. H. Hwang. S.W Shinn, Retailer's pricing and lot sizing policy for exponentially deteriorating

- products under the condition of permissible delay in payments. *Comput. Oper. Res.* 24(1937) 539-547.
8. B.R. Sarkar, A.M.M Jamal, S. Wang, Supply chain model for perishable products under inflation and permissible delay in payment *Comput. Oper. Res.* 27(2000) 59-75.
  9. N.H. Shah, A lot size model for exponentially decaying inventory when delay in payments in permissible *Coh. Du (ERO)* 35(1993) 115-123.
  10. S.W. Shinn, Determining optimal retail and lot size under day terms supplier credit, *comput. Ind. Eng.* 33(1997) 717-720.
  11. M.Khouja, A. Mehrez, Optimal inventory policy under different supplier credits, *J. Manuf. Syst.* 15 (1996) 334-339.
  12. Chung K.J. and Ting P.S., "A heuristic for replenishment of deteriorated items with a linear trend in demand." *Journal of the operational research society*, 44(1993) 1235-1241.
  13. Dave, V. and Patel L.K "Order level inventory system with power demand pattern for items with variable rate of deterioration". *Indian journal of pune and applied mathematics*, 19(1988) 1043-1053.
  14. Deb, M. and Chaudhari , K.S. "An EOQ model for items with finite rate of production and variable rate of deterioration," *Journal of the operational research society*, 43(1992) 993-1001.
  15. Donaldson, W.A. "Inventory replenishment policy for a linear trend in demand: an analytic solution", *operational research quarterly*, 28(1977) 663-670.
  16. Giri B.C, Goswami A., and Chaudhari, K.S., "An EOQ model for deteriorating items with time varying demand and costs." *Journal of operational research society*, 47(1996) 1398-1405.
  17. Goel, V.P, and Aggrawal, S.P. , "Oprder level inventory system with power demand pattern for deteriorating items." *Proceedings of All India Seminar on Operational Research and Decision Making*, University of Delhi, Delhi, 198.
  18. Hwang, H.S. "A study on an Inventory model for items with weibull ameliorating", *com and Ind. Eng.* 33(1997) 701-704.
  19. Kishan, H., and Mishra, P.N. "An inventory model with exponential demand and constant deterioration with shortages", *Indian Journal of Mathematics*, 37(1995) 275-279.
  20. Pal. A.K, and Mandal, B. "An EOQ model for deteriorating inventory with alternating demand rates". *The Korean Journal of Computational and Applied Mathematics*, 4(1997) 397-407.
  21. Ritchie, E., "The EOQ for linear increasing demand, a simple solution", *Journal of the operational research society* 35(1984) 949-952.



# Significant Role of Statistics in Computational Sciences

Rakesh Kumar Singh  
Scientist-D  
G.B. Pant Institute of  
Himalayan Environment &  
Development  
Kosi-Katarmal, Almora  
Uttarakhand, India

Neeraj Tiwari  
Professor & Head  
Department of Statistics  
Kumaun University  
SSJ Campus, Almora  
Uttarakhand, India

R.C. Prasad  
Scientist-F  
G.B. Pant Institute of  
Himalayan Environment &  
Development  
Kosi-Katarmal, Almora  
Uttarakhand, India

**Abstract:** This paper is focused on the issues related to optimizing statistical approaches in the emerging fields of Computer Science and Information Technology. More emphasis has been given on the role of statistical techniques in modern data mining. Statistics is the science of learning from data and of measuring, controlling, and communicating uncertainty. Statistical approaches can play a vital role for providing significance contribution in the field of software engineering, neural network, data mining, bioinformatics and other allied fields. Statistical techniques not only helps make scientific models but it quantifies the reliability, reproducibility and general uncertainty associated with these models. In the current scenario, large amount of data is automatically recorded with computers and managed with the data base management systems (DBMS) for storage and fast retrieval purpose. The practice of examining large pre-existing databases in order to generate new information is known as data mining. Presently, data mining has attracted substantial attention in the research and commercial arena which involves applications of a variety of statistical techniques. Twenty years ago mostly data was collected manually and the data set was in simple form but in present time, there have been considerable changes in the nature of data. Statistical techniques and computer applications can be utilized to obtain maximum information with the fewest possible measurements to reduce the cost of data collection.

**Keywords:** Statistics, Data Mining, Software Engineering, DBMS, Neural Networks, etc

## 1. INTRODUCTION

Statistics is a scientific discipline having sophisticated methods for statistical inference, prediction, quantification of uncertainty and experimental design. From ancient to modern times statistics has been fundamental to advances in computer science. The statistics encompasses a wide range of research areas. The future of the World Wide Web (www) will depend on the development of many new statistical ideas and algorithms. The most productive approach is involve with statistics are: computational and mathematical. Modern statistics encompasses the collection, presentation and characterization of information to assist in both data analysis and the decision-making process. Statistical advances made in collaboration with other sciences can address various challenges in the field of science and technology. Computer science uses statistics in many ways to guarantee products available on the market are accurate, reliable, and helpful[1][2].

- **Statistical Computing:** The term “statistical computing” to refer to the computational methods that enable statistical methods. Statistical computing includes numerical analysis, database methodology, computer graphics, software engineering and the computer-human interface[1].
- **Computational Statistics:** The term “computational statistics” somewhat more broadly to include not only the methods of statistical computing but also modern statistical methods that are computationally intensive. Thus, to some extent, “computational statistics” refers to a large class of modern statistical methods. Computational statistics is grounded in mathematical statistics, statistical computing and applied statistics. Computational statistics is related to the advance of statistical theory and methods

through the use of computational methods. Computation in statistics is based on algorithms which originate in numerical mathematics or in computer science. The group of algorithms highly relevant for computational statistics from computer science is machine learning, artificial intelligence (AI), and knowledge discovery in data bases or data mining. These developments have given rise to a new research area on the borderline between statistics and computer science[1].

- **Computer Science vs. Statistics:** Statistics and Computer Science are both about data. Massive amounts of data is present around today’s World. Statistics lets us summarize and understand it with the use of Computer Science. Statistics also lets data do our work for us[2].

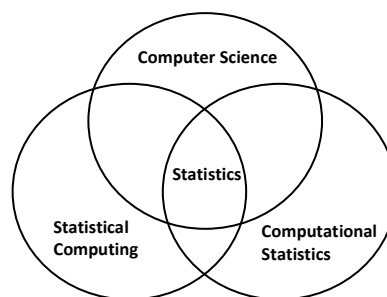


Fig.1. Relation between statistics, computer science, statistical computing and computational statistics.

## 2. STATISTICAL APPROACHES IN COMPUTATIONAL SCIENCES

Statistics is essential to the field of computer science in ensuring effectiveness, efficiency, reliability, and high-quality

products for the public. Statistical thinking not only helps make scientific discoveries, but it quantifies the reliability, reproducibility and general uncertainty associated with these discoveries. The following terms are a brief listing of areas in computer science that use statistics to varying degrees at various times[6][7][8]:

- **Data Mining:** Data mining is the analysis of information in a database, using tools that look for trends or irregularities in large data sets. In other words "finding useful information from the available data sets using statistical techniques".
- **Data Compression:** Data compression is the coding of data using compact formulas, called algorithms, and utilities to save storage space or transmission time.
- **Speech Recognition:** Speech recognition is the identification of spoken words by a machine. The spoken words are turned into a sequence of numbers and matched against coded dictionaries.
- **Vision and Image Analyses:** Vision and image analyses use statistics to solve contemporary and practical problems in computer vision, image processing, and artificial intelligence.
- **Human/Computer Interaction:** Human/Computer interaction uses statistics to design, implement, and evaluate new technologies that are useable, useful, and appealing to a broad cross-section of people.
- **Network/Traffic Modeling:** Network/Traffic modeling uses statistics to avoid network congestion while fully exploiting the available bandwidth.
- **Stochastic Optimization:** Stochastic optimization uses chance and probability models to develop the most efficient code for finding the solution to a problem.
- **Stochastic Algorithms:** Stochastic algorithms follow a detailed sequence of actions to perform or accomplish a task in the face of uncertainty.
- **Artificial Intelligence:** Artificial intelligence is concerned with modelling aspects of human thought on computers.
- **Machine Learning:** Machine learning is the ability of a machine or system to improve its performance based on previous results.
- **Capacity Planning:** Capacity planning determines what equipment and software will be sufficient while providing the most power for the least cost.
- **Storage and Retrieval:** Storage and retrieval techniques rely on statistics to ensure computerized data is kept and recovered efficiently and reliably.
- **Quality Management:** Quality management uses statistics to analyze the condition of manufactured parts (hardware, software, etc.) using tools and sampling to ensure a minimum level of defects.
- **Software Engineering:** Software engineering is a systematic approach to the analysis, design, implementation, and maintenance of computer programs.
- **Performance Evaluation:** Performance evaluation is the process of examining a system or system component to determine the extent to which specified properties are present.
- **Hardware Manufacturing:** Hardware manufacturing is the creation of the physical material parts of a system, such as the monitor or disk drive.

### 3. STATISTICS IN SOFTWARE ENGINEERING

Software engineering aims to develop methodologies and procedures to control the whole software development process. Nowadays researchers attempt to bridge the islands of knowledge and experience between statistics and software engineering by enunciating a new interdisciplinary field: *statistical software engineering*. Design of Experiments (DOE) uses statistical techniques to test and construct models of engineering components and systems. Quality control and process control use statistics as a tool to manage conformance to specifications of manufacturing processes and their products. Time and methods engineering uses statistics to study repetitive operations in manufacturing in order to set standards and find optimum (in some sense) manufacturing procedures. Reliability engineering uses statistics to measure the ability of a system to perform for its intended function (and time) and has tools for improving performance. Probabilistic design uses statistics in the use of probability in product and system design. Essential to statistical software engineering, is the role of data: *wherever data are used or can be generated in the software life cycle, statistical methods can be brought to bear for description, estimation, and prediction*. The department of software engineering and statistics trains multiskilled engineers in the processing of information, both in its statistical and computational forms, for use in various business professions.

### 4. STATISTICS IN HARDWARE MANUFACTURING

The hardware manufacturing companies are applying statistical approaches to create a plan of action that will work more efficiently for forecasting the future productivity of the hardware enterprise[8]. Adopted statistical approaches for:

- Forecasting production, when there is a stable demand and uncertain demand.
- Pinpoint when and which inputs of a specific model will be the cause of uncertainty
- Calculate summary statistics in order to set sample data.
- To make market analysis and process optimizations.
- Statistical tracking and predicting for quality improvement

### 5. STATISTICS IN DATABASE MANAGEMENT

Databases are packages designed to create, edit, manipulate and analyze data. To be suitable for a database, the data must consist of records which provide information on individual cases, people, places, features, etc. Optimizer statistics are a collection of data that describe more details about the database and the objects in the database. The optimizer statistics are stored in the data dictionary. They can be viewed using data dictionary views. Because the objects in a database can be constantly changing; statistics must be regularly updated so that they accurately describe these database objects. These statistics are used by the query optimizer to choose the best execution plan for each SQL statement[5]. Optimizer statistics include the following:

- Table Statistics
  - Number of rows
  - Number of blocks
  - Average row length
- Column Statistics

- Number of distinct values (NDV) in column
- Number of nulls in column
- Data distribution (histogram)
- Index Statistics
  - Number of leaf blocks
  - Levels
  - Clustering factor
- System Statistics
  - I/O performance and utilization
  - CPU performance and utilization

Statistical packages for databases are SAS, SPSS, R, etc. and these are available over a wide range of operating systems. Numerous other packages have been developed specifically for the PC DOS environment. S is a commonly available statistical package for UNIX

## 6. STATISTICS IN ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is the intelligence exhibited by machines or software. Popular AI approaches include statistical methods, computational intelligence, machine learning and traditional symbolic AI. The goals of AI include reasoning, knowledge, planning, learning, natural language processing, perception and the ability to move and manipulate objects. There are a large number of tools used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics, and many others[4]. The simplest AI applications can be divided into two types:

- **Classifiers:** Classifiers are functions that use pattern matching to determine a closest match. A classifier can be trained in various ways; there are many statistical and machine learning approaches. The most widely used classifiers is the neural network.
- **Controllers:** Controllers do however also classify conditions before inferring actions, and therefore classification forms a central part of many AI systems.

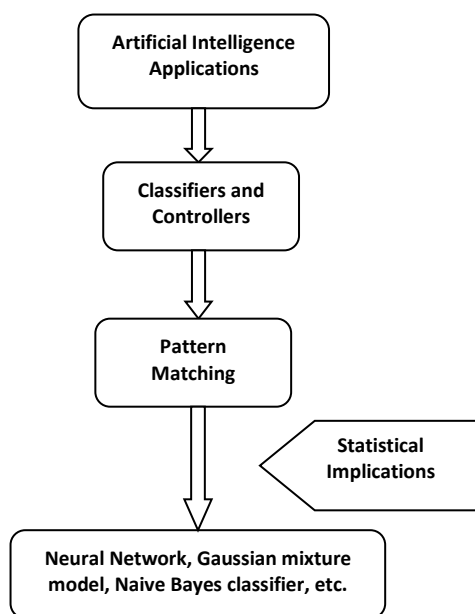


Fig.2. Graphical approach of Artificial Intelligence.

## 7. STATISTICS IN NEURAL NETWORK

Neural network had been used to refer to a network of biological neurons and artificial neural networks used to refer to a network of artificial neurons or nodes. Biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. Artificial neural networks are made up of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons). Artificial neural networks may either be used to gain an understanding of biological neural networks or for solving artificial intelligence problems without necessarily creating a model of a real biological system. Because the inner product is a linear operator in the input space, the Perception can only perfectly classify a set of data for which different classes are linearly separable in the input space, while it often fails completely for non-separable data. While the development of the algorithm initially generated some enthusiasm, partly because of its apparent relation to biological mechanisms, the later discovery of this inadequacy caused such models to be abandoned until the introduction of non-linear models into the field[4].

## 8. STATISTICS IN BIOINFORMATICS

Bioinformatics is the application of "computational biology" to the management and analysis of biological data. Concepts from computer science, discrete mathematics and statics are being used increasingly to study and describe biological systems. Bioinformatics would not be possible without advances in computer hardware and software: analysis of algorithms, data structures and software engineering. To elaborate algorithms on computers increased the awareness of more recent statistical methods. Statistical analysis for differently expressed genes are best carried out via hypothesis test. More complex data may require analysis via ANOVA or general linear models[8].

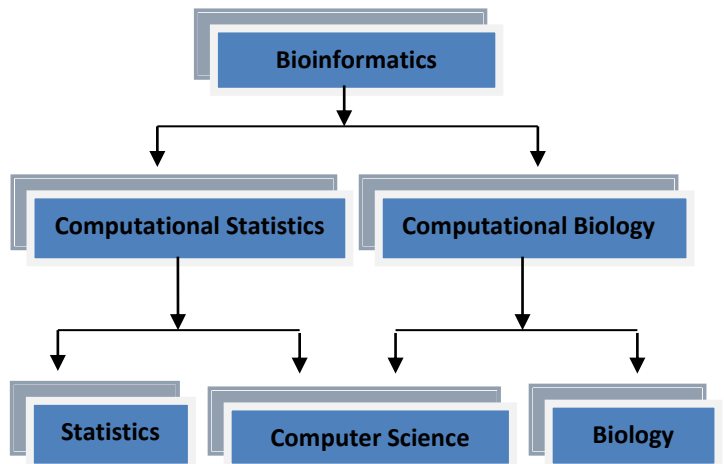


Fig.3. Taxonomy of Bioinformatics.

## 9. STATISTICS IN DATA MINING

Data Mining is a process of discovering previously unknown and potentially useful hidden pattern in the data. Advances in information technology have resulted in a much more data-based society. Data touch almost every aspect of our lives like commerce on the web, measuring our fitness and safety, doctors treat our illnesses, economic decisions that affect entire nations, etc. Alone, data are not useful for knowledge

discovery. Data mining are transitioning from data-poor to data-rich by using the methods like data exploration, statistical inference and understanding of variability and uncertainty[5].

**Statistical Elements Present in Data Mining**

- Contrived serendipity, creating the conditions for fortuitous discovery.
- Exploratory data analysis with large data sets, in which the data are as far as possible allowed to speak for themselves, independently of subject area assumptions and of models which might explain their pattern. There is a particular focus on the search for unusual or interesting features.
- Specialised problems: fraud detection.
- The search for specific known patterns.
- Standard statistical analysis problems with large data sets.

**Data Mining from Statistical Perspective**

- Data sets which are relatively large and homogeneous might be reasonable to us mainstream statistical techniques on the whole or a very large subset of the data.
- All analyses done by mainstream statistics have intended outcome like set of data to a small amount of readily assimilated information.
- The outcome may include graphs, or summary statistics, or equations that can be used for prediction or a decision tree.
- Large volume of data without loss of information be reduced to a much smaller summary form, this can enormously aid the subsequent analysis task.
- It becomes much easier to make graphical and other checks that give the analyst assurance that predictive models or other analysis outcomes are meaningful and valid

**Statistics vs. Data Mining**

| Feature                                       | Statistics  | Data Mining  |
|---|---|--|
| Type of Problem                               | Well structured   | Unstructured / Semi-structured                                     |
| Inference Role                                | Explicit inference plays great role in any analysis       | No explicit inference  |
| Objective of the Analysis and Data Collection | First – objective formulation, and then - data collection | Data rarely collected for objective of the analysis/modeling       |
| Size of data set                              | Data set is small and hopefully homogeneous               | Data set is large and data set is heterogeneous                    |
| Paradigm/A pproach                            | Theory-based (deductive)                                  | Synergy of theory-based and heuristic-based approaches (inductive) |
| Type of Analysis                              | Confirmative  | Explorative  |
| Number of variables                           | Small   | Large  |

| Methods/Techniques | - Dependence Methods: Discriminant analysis, Logistic regression<br>- Interdependence Methods: Correlation analysis, Correspondence analysis, Cluster analysis | - Predictive Data Mining: Classification, Regression<br>- Discovery Data Mining: Association Analysis, Sequence Analysis, Clustering |
|--------------------|--|--|
|                    |  |  |

**10. PROPERTIES OF STATISTICAL PACKAGES**

Statistical packages offer a range of types of statistical analysis[3]. Statistical packages includes:

- Database functions, such as editing, printing reports.
- Capabilities for graphic output, particularly graphs but many also produce maps.
- Common packages are SAS, SPSS, R, etc.
- Available over a wide range of operating systems.
- Some have been "ported" to (rewritten for) the IBM PC.
- Numerous other packages have been developed specifically for the PC DOS environment.
- S is a commonly available statistical package for UNIX

**11. CONCLUSION**

In this paper, many areas of computer science have been described in which statistics plays a very vital role for data and information management. Statistical thinking fuels the cross-fertilization of ideas between scientific fields (biological, physical, and social sciences), industry, and government. The statistical and algorithmic issues are both important in the context of data mining. Statistics is an essential and valuable component for any data mining exercise. The future success of data mining will depend critically on our ability to integrate techniques for modeling and inference from statistics into the mainstream of data mining practice.

**12. REFERENCES**

- [1] Lauro, C. (1996). Computational Statistics or Statistical Computing, is that the question? *Computational Statistics and Data Analysis*, Vol. 23, pp.191–193.
- [2] Billard, L. and Gentle, J.E. (1993). The middle years of the Interface. *Computing Science and Statistics*, Vol. 25, pp.19–26.
- [3] Yates, F (1966). Computers: the second revolution in statistics. *Biometrics*, Vol. 22.
- [4] Cheng, B. and Titterington, D. M. (1994). Neural networks: a review from a statistical perspective. *Statistical Science*, Vol. 9, pp.2-54.
- [5] Elder, J. F. and Pregibon, D. (1996). A statistical perspective on knowledge discovery in databases. *Advances in Knowledge Discovery and Data Mining*, MIT Press, pp.83-115.
- [6] Gentle, J.E. (2004). Courses in statistical computing and computational statistics. *The American Statistician*, Vol. 58, pp.2–5.
- [7] Grier, D.A. (1991). Statistics and the introduction of digital computers. *Chance*, Vol. 4(3), pp.30–36.
- [8] Friedman, J. H. and Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, Vol. 9, pp.123-143.



# SPHMS : Smart Patient m-Healthcare Monitoring System with NFC and IOT

Hoda Ramin Hossein  
Computer Department,  
University of Pune, AISSMS,  
COE  
Pune, Maharashtra 411001,  
India

---

**Abstract:** In the latest technology, NFC(Near Field Communication) is a recently growing technology for communication for short range which is aimed to augment existing near field technologies such as RFID(Radio Frequency Identification). In this paper, we have to propose a system which describes how an android application is used as our own NFC tag writer to write patient unique id in NFC tag. Doctors using NFC enabled smart phones to retrieve patient information when placed near NFC tag. This proposed system uses an android smart phone which automatically monitors the patients status and keeps the doctor/nurse updated and maintained through logs that are stored on a cloud server and also alerts are generated through /Email in case of emergency/detection of death of any patient. This system helps in reducing the paper work required during the registration of the patient and also eliminates the disadvantages of using RFID based system.

**Keywords:** NFC tags, web server, NFC enabled smart phone, Android application.

---

## 1. INTRODUCTION

An important requirement now-a-days for medical service informalization is the establishment and improvement of doctor-patient interaction system. Especially in today's era where the mobile communication technology develops rapidly, it is a challenge to make the mobile web of full use to make up for the time and distance gap between doctors and patients to provide fast and adequate medical services or not becomes an important factor to measure hospital's competitive ability. Hospitals and clinics are increasingly turning to modern technology to facilitate treatment and optimize patient management

In current technology enabled world, there is rapid change and the status-quo is constantly disrupted. Internet Of Things (IOT) is one such field where disruption is taking place right now and has the potential to change the way healthcare is delivered. IOT does not have any standard definition, as per Gartner, "IOT is a network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment". People and things can be easily connected to each other at anytime, anyplace with the help of using any path or network and any service with the help of IOT.

There is one other technology emerging now-a-days which is Near Field Communication (NFC). NFC is known to be a wireless technology used for communication in a short range of four to ten centimetres. NFC is based on RFID technology which by using a reading device can be used to read information from a tag. By working in three different modes NFC provides excellent security and privacy. These modes are Read/write mode, Peer to peer mode, Card emulation mode.

### 1. Card Emulation mode:

An NFC device in a card emulation mode can replace a contactless card or a tag. This will enable NFC devices to be used with existing contactless card infrastructure in applications such as access control, payments and ticketing.

### 2. Reader/Writer mode:

An NFC enabled device is able to read and write to NFC and many contactless cards. For example, if a NFC tag is attached to a poster, the NFC Smartphone can "tap" the tag to access the information stored in the tag easy and conveniently.

3. Peer to Peer mode: A Peer to peer mode enables two NFC devices to share data between them. Here, the NFC is used to negotiate the optimal communication protocol and transfer authentication data for secondary protocol. Then the file or data is sent over the high capacity protocol(i.e Bluetooth, Ei-Fi etc).

In this paper, we have proposed a system which describes how an android application is used as our own NFC tag writer to write patient unique id in NFC tag. This will help in reducing the paper work that is required during the registration when a patient is admitted in the hospital. Doctors will be using NFC enabled smart phones to retrieve patient information when placed near NFC tag. This system uses a smart phone which automatically monitors the patients status and keeps the doctor/nurse updated and maintained through logs that are stored on a cloud server and also alerts are generated through SMS/Email in case of emergency/detection of death of any patient. This system helps in reducing the paper work required during the registration of the patient and also eliminates the disadvantages of using RFID based system.



## 2. RELATED WORK

### 2.1 Literature Survey

The existing system in healthcare involves manual intervention where there is note taking, updating the notes to the computer and maintaining the records under a unique id assigned to every patient. This process is usually very slow and error prone due to which there is latency between data gathering and information accessibility.

“PSMPA: Patient Self-controllable & Multi-level Privacy Preserving cooperative authentication in distributed m-healthcare cloud computing system”[1] has introduced security in m-Healthcare by implementing a three tier architecture and providing security on the cloud where all the data of the patient is stored.

In recent work, “Design and Implementation of Doctor-Patient system based on android”[2] introduced m-health and Android platform which was designed to achieve a set of doctor-patient interaction system based on android. One other system A wireless ECG monitoring system[2] was implemented which integrated novel capacitive ECG sensors.

NFC is playing an important role, in “NFC-based hospital Real-Time Patient management system”[8], they have used NFC technology where NFC smart tag or wrist band is given to each patient during the time of registration. Then NFC encoded devices can be used to read the information from the patient without causing any inconvenience at any time.

In other work[9], they have used an android application to view the patients data with the help of an open source platform. This system has helped the doctor’s in monitoring and viewing the patient’s data through a mobile device.

### 2.2 Existing System

In our real time system, manual intervention is required i.e when a patient is admitted the registration of the patient is done manually which is more tedious and error prone job. Also when a patient is in critical condition, no alerts are generated which may sometimes cause loss of life. Also doctors don’t have a provision of remote monitoring.

Before NFC, the similar device used was RFID. With using RFID the cost was high as there was a requirement of RFID reader. Also the architecture used was servlets technology which was not flexible as the clients of only java technology could be interfaced.

Taking all this drawbacks of existing system we have introduced the concept of NFC and cloud server. By using NFC tags, the cost of using RFID reader is cut down. And also by using cloud server i.e web server we can interface clients of any technology with the server.

## 3. PROPOSED SYSTEM

### 3.1 Problem Statement

“To develop an automated system using smart phone which will automatically monitor patients status and generate alerts and maintain logs on cloud server.”

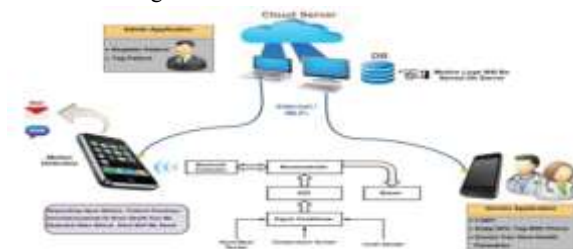


Figure 1: Proposed System Architecture

### 3.2 Proposed Architecture

The proposed architecture consist of various application Viz Admin application, doctor application, android application and cloud server. There is a hardware device connected to the android application through Bluetooth. The hardware device consist of sensors connected to it. In this paper we will be using basically 3 sensors i.e heart rate sensor, body temperature sensor and level sensor [saline level].

#### 1. Patient Position Sensor :

The Patient Position Sensor (Accelerometer) monitors five different patient positions (prone, supine, standing/sitting ,left and right.) In many cases, it is necessary to monitor the body positions and movements made because of their relationships to particular diseases (i.e., sleep apnea and restless legs syndrome). The body position sensor could help also to detect fainting or falling of elderly people or persons with disabilities and also detect death status through a motion detector.

#### 2. Body Temperature Sensor :

The commonly accepted average core body temperature (taken internally) is 37.0°C (98.6°F). In healthy adults, body temperature fluctuates about 0.5°C (0.9°F) throughout the day, with temperatures being lower in the morning and higher in late afternoon and evening depending on the body's needs and activities the temperature change. Measuring body temperature is of great medical importance. The reason is that a by characteristic change in body temperature a number of diseases are accompanied. Likewise, by measuring body temperature, the course of certain diseases can be monitored, and the efficiency of a treatment initiated can be evaluated by the physician.

#### 3. Heart Rate Sensor :

A normal resting heart rate for adult’s ranges from 60 to 100 beats a minute. Generally, a lower heart rate at rest implies more efficient heart function and better cardiovascular fitness. The Heart Rate speeds up or slows down in a number of ways Bradycardia is defined as a resting heart rate below 60 bpm.

However, heart rate from 50 to 60 bpm are common among healthy people and do not necessarily require special attention. Tachycardia is defined as a resting heart rate above 100 bpm, though persistent rest rates between 80-100 bpm, mainly if they are present during sleep, may be signs of hyperthyroidism or anemia.

The number of sensors can be increased upto eight. The hardware device consist of a number of components i.e Bluetooth controller, microcontroller, ADC(Analog to Digital), signal conditioner and buzzer. The sensors are connected to the signal conditioner through which all the values of the sensors are collected and then sent to the ADC where the signal is converted from analog to digital and then sent to the microcontroller which transmits the data to the android application through a Bluetooth controller. In case of emergency the buzzer will ring or an SMS/Email will be sent to the doctor on the smart phone.

All the data and information related to the patient is stored on the cloud server. We have introduced cloud server so that clients of different platform can be interfaced which overcomes the disadvantage of using servlets.

Patient will be monitored for unconsciousness or death status through the phone camera. The state of the patient will be recorded and sent after a fixed time interval and in case of any emergency the buzzer will ring and also alerts will be sent through SMS or Email.

NFC tags will be used with the help of an NFC enabled smart phone by the doctor to retrieve the patient's information. By using NFC tags we are eliminating the paper work that had to be done during the registration of the patient when he/she is admitted

#### **4. PROPOSED ARCHITECTURE IMPLEMENTATION**

When a patient is admitted in hospital for the first time a unique id is provided to patient. NFC tag will be associated with every patient. equipped with NFC enabled smart phones will be equipped to Doctors and other staff. There will be an NFC issue app which will be used to read the content from mobile to NFC tag. Administrator can write unique tag id and application link in NFC tag by using the NFC issue app. To create patient application link administrator use IP of the server and then transfer to mobile by using same IP. Whenever NFC tag is placed near NFC enable smartphones the patient data is retrieved directly from the server. Doctor can also do the same thing to retrieve information of the patient from the server.

After the patient is admitted, android phone will be associated which will have motion detection to detect the death/unconsciousness of the patient. Also there will be different sensors that will monitor different parameters such as heart beat rate, body temperature and level sensor (saline level). after a time interval the data will be transmitted through Bluetooth to the android phone which will further be sent to the cloud server through WiFi and stored there. The doctor

can view the data any time and also can be viewed in graphical form.

#### **5. CONCLUSION**

In this proposed system, we have overcome the disadvantages of using RFID system by using NFC tags. We have also replaced the servlets architecture by using web service architecture. Hence the patients status is automatically generated and stored in the cloud server and also alerts are generated if the patient is in a critical condition. Currently it is possible to conclude that for providing a better solution to the m-healthcare hospital management system in many hospitals NFC promises appear to be credible.

#### **6. ACKNOWLEDGMENTS**

I am thankful to my guide Mrs. S.S.Shaikh for her complete support and guidance throughout the work.

#### **7. REFERENCES**

- [1] J. Zhou, Xiaodong Lin, et.al. "PSMPA: Patient Self-controllable & Multi-level Privacy Preserving cooperative authentication in distributed m-healthcare cloud computing system" in IEEE transaction on parallel and distributed systems, vol. 26, No. 6, June 2015, Pg. 1693-1703..
- [2] Ran Wei, "Zhimin Yang", "Design and Implementation of Doctor-Patient system based on android" in 2012 international symposium on information technology in medicine and education(2012), Pg. 580-583.
- [3] Ying-Wei Bai, Wen-Tai and You-Wei Chen, "Design and implementation of an embedded monitor system for detection of a patient's breath by double webcams" in medical measurements & applications proceedings(MeMeA), 2010 IEEE international workshop on april 30-may 1 2010, Pg. 171-176.Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
- [4] J. Mirkovic, H. Bryhmi and C.M. Ruland, "A framework for the development of Ubiquitous patient support system", in pervasive health 2012, May 21-24, San Diego, United States, Pg. 81-88.
- [5] Ms. Sohate R. Pawara, "Instant notification system in heterogeneous network with deployment of XMPP protocol" in Cloud & Ubiquitous Computing & Emerging Technologies(CUBE), 2013 international conference on 15-16 Nov, 2013, Pg. 87-92.
- [6] P. Khanja, S. Wattanasirichaigan, Et.Al., "A web base system for ECG data Transferred using Zigbee/IEEE technology" in the 3<sup>rd</sup> international symposium on Biomedical engineering(ISBME 2008) Pg. 109-112.
- [7] Mehmet R., Yuce Peng Choong Ng and Jamil Y. Khan "Monitoring of Physiological parameters from multiple patients using wireless sensor network" in Journal of Medical systems, vol. 32, IS. 5,PP 433-441
- [8] AV.Gopi Krishna, C.Sreevardhan, S.Karun, S.P.Kumar, "NFC-based hospital Real-Time Patient management system" in IJETT- Vol.4 IS 4- April 2013 Pg. 626-629.

- [9] S.Challa, G.Geethakumari, CSN Prasad, "Patient Data Viewer: An Android application for healthcare" in India Conference(INDICON), 2011 Annual IEEE, 16-18 Dec 2011.
- [10] Lei Clifton, D.A.Clifton, M.A.F.Pimentel, Peter J. Watkinson and L. Tarassenko, "Predictive Monitoring of Mobile patients by combining clinical observations with Data from wearable sensors" in IEEE Journal of Biomedical and Health Informatics, Vol. 18, No 3, May 2014 Pg. 722-730.

# Destination Aware APU Strategy for Geographic Routing in MANET

Yukti

M.tech Student

Department of Computer science and engineering  
Giani Zail Singh- Punjab Technical University Campus  
Bathinda, Punjab, India

Abhilasha jain

Associate professor

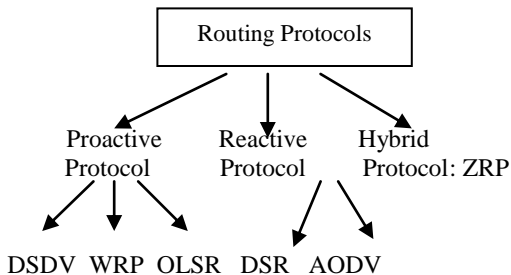
Department of Computer science and engineering  
Giani Zail Singh- Punjab Technical University Campus  
Bathinda, Punjab, India

**Abstract:** In this paper, we have explained the Enhanced Adaptive Position Update strategy for geographic routing in mobile ad hoc network. In Adaptive Position Update strategy, there are two techniques: Mobility prediction rule and On-demand learning rule. Proposed system is based on the destination aware routing in which path to transfer the data over the network is based on the distance from highly stable node to the destination node. Results of the proposed system are compared with Periodic Beaconing on the basis of packet delivery ratio, beacon overhead, energy consumption. Experiment results show a high improvement in results on the parameters energy consumption, packet delivery ratio and beacon overhead. Proposed work is implemented on the NS2 (Network Simulator) Environment to perform experiments.

**Keywords:** APU, Enhanced APU, MANET, Destination aware routing.

## 1. INTRODUCTION

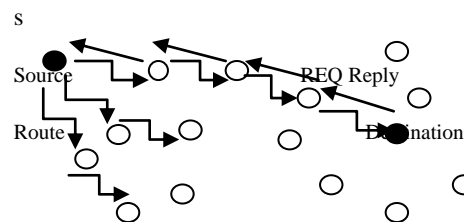
MANET is self-configured Mobile Ad hoc network of mobile nodes connected via random topology. The main issue in MANET is the selection of best path between any two nodes [1]. Routing in MANET is to find the efficient path to transfer the data from source to destination.



**Figure 1. Categorization of routing protocols**

Main protocols used for routing in MANET are categorized as Proactive protocols, Reactive protocols and hybrid protocols. Proactive protocols are those which are derived from past study and information is maintained in the form of tables so also called table-driven (eg. DSDV, WRP, OLSR). In Reactive protocols, instead of past approaches the nodes send the data packets to the destination on-demand (eg. DSR, AODV). In hybrid protocols, the combination of both the reactive and proactive protocols, this protocol includes ZRP (Zone Routing Protocol) [2]. These protocols are briefly explained as: In DSDV, enhancement is done on Bellman-ford Routing Algorithm by making it loop-free. In this, each node maintain a routing table which contains the entries of all the devices in network and table is updated via periodically broadcast routing messages to its neighbor. In WRP, which also resembles the DSDV protocol and its main idea is to find the shortest distance to every node and it is basically a path-finding algorithm with

exception that it counters the count-to-infinity problem by forcing each node to perform consistency checks of predecessor information reported by all its neighbors. This eliminates looping situations and enables faster route convergence when a link failure occurs. In OLSR, point-to-point approach that employs a link-state mechanism known as multipoint relaying. Two ways are there: by reducing the size of control packets or by reducing the number of links used in forwarding the packets. The methodology used is first the analysis of changes in the node's neighborhood called neighbor sensing then find the best path to every destination using this information and store it in table called routing table. Then the best route for destination is computed via shortest-path algorithm. In DSR fig.2, the unique feature is that it doesn't use periodic updates. In this, basically two things happen: route discovery and route maintenance. First route is discovered on-demand i.e. it finds the route when necessary for that sender determines the complete series of the nodes from source to destination and lists this route in packet header and save it in cache. If source wants to send data it first check the cache if there is entry in cache for that destination, I transfer the packet otherwise it broadcast a route request to all its neighbors and wait, then the node having the route information send the request reply to source and sender send the packets to discovered route and also make entry in cache for further use.



**Figure 2. DSR protocol**

In AODV, the concept of DSR is used in addition; it uses the hop-to-hop routing, sequence numbers and periodic beacons from DSDV. It removes the drawback of WRP proactive protocol and finds the route only if required. The main problems in AODV are: i) gradual increase in sequence numbers and ii) decrease in hop count and It is not secure. In ZRP, advantages of both proactive and reactive protocols are mixed. It works in two parts of which first uses the table-driven approach to speed up communication and later uses the on-demand approach to avoid unnecessary communication [3].

Further comes the concept of geographic routing which is widely used in Mobile ad hoc networks due to its simplicity and scalability and has low overhead compared to other protocols. Geographic routing based on assumption that every node is aware of its location in the network (location information generated via GPS and schemes explained in [16]). Based on this assumption, various protocols are derived and worked on: DREAM, LAR, GPSR. These routing protocols are explained as follows: In DREAM, nodes forwards data packets to all the neighbors that lie in direction of destination and every node updates every other node about its location. In LAR, it is assumed that every node is aware of its location but can't predict location of other nodes. It basically reduces the flooding overhead. Two LAR schemes are proposed is sending node is only within request zone and outer area is not addressed and second is the nodes always send the packets to node closer to the destination [5]. In GPSR, nodes forwards the data packets based on distance. It comprises two techniques: greedy forwarding and face routing. Till the destination reached, the forwarding is carried out on greedy basis by selecting the closest node to destination. If this method doesn't work then face routing is used. After the node is in transmission range then again greedy forwarding is used [6].

The remaining part of the paper is organized as: In Section 2, we discuss related work or the literature survey then in Section 3 the proposed scheme and Section 4 signifies the results and Section 5 concludes the paper.

## 2. RELATED WORK (LITERATURE SURVEY)

Many papers and articles are proposed on Mobile ad hoc Network and the protocols defining the routing of packets among network. There are Distance routing protocols and Location based routing protocols and further we are using new technique for the same. According to [1], GeRaF is the forwarding technique based on the assumption that nodes has a medium to detect their location and position of final destination and transmitting node is included in the forwarding message.

### Problems with the existing system

Problems in Position updates in network are costly in many ways.

1. Each update in Position in a given network consumes node energy, bandwidth of wireless channel, and the

risk of packet collision increases correspondingly at the medium access control (MAC) layer.

2. Packet collisions in the network cause packet loss which again decreases the routing performance which results in difficulty to evaluate local topology for transmission. It results in a heavy loss due this factor beacon lost and a lost beacon cannot be retransmitted.
3. A lost data packet does get retransmitted as a result end-to-end delay increases.

In the existing system, Periodical beacon broadcasting is used to exchange the neighbor's location that results in degradation in performance, increases packet loss and results in longer end to end delay. Hence a more robust update policy is required to be discovered to overcome all these types of the costs.

## 3. PROPOSED WORK

An Enhanced APU strategy in the proposed system use destination aware routing with minimum distance dynamically adjusts the beacon update intervals based on the minimum distance from source to destination, *mobility dynamics* of the nodes and the *forwarding patterns* in the network.

The beacons transmitted by the nodes contain their current position and speed. Nodes estimate their positions periodically by employing linear kinematic equations based on the parameters announced in the last announced beacon. If the predicted location is different from the actual location, a new beacon is broadcast to inform the neighbors about changes in the node's mobility characteristics.

An accurate representation of the local topology is particularly desired at those nodes that are responsible for forwarding packets. Hence, APU seeks to increase the frequency of beacon updates at those nodes that overhear data packet transmissions. As a result, nodes involved in forwarding packets can build an enriched view of the local topology.

### Algorithm for selection of forwarder

- Step1:** Find distance  $[d(t)]$  of each neighbor from source at time  $T$
- Step2:** Find distance  $[d(t+T)]$  of each neighbor from source at time  $(T+t)$
- Step3:** If  $\{ [d(t+T)] - [d(t)] < \text{Threshold} \} \rightarrow$  Select the neighbor as high stable link
- Step4:** Find distance  $D_{des}$  between destination and the node having high stable link
- Step5:** Link having minimum  $D_{des}$  is selected as next hop

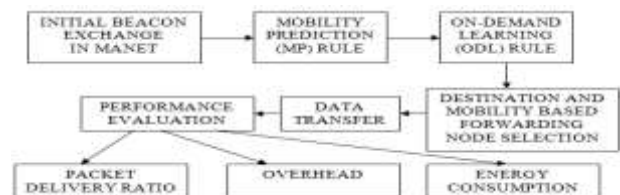


Figure 3. Block Diagram representing the functions in the proposed APU system



In the above fig.3 different modules of the proposed APU system to update the position of beacon are represented.

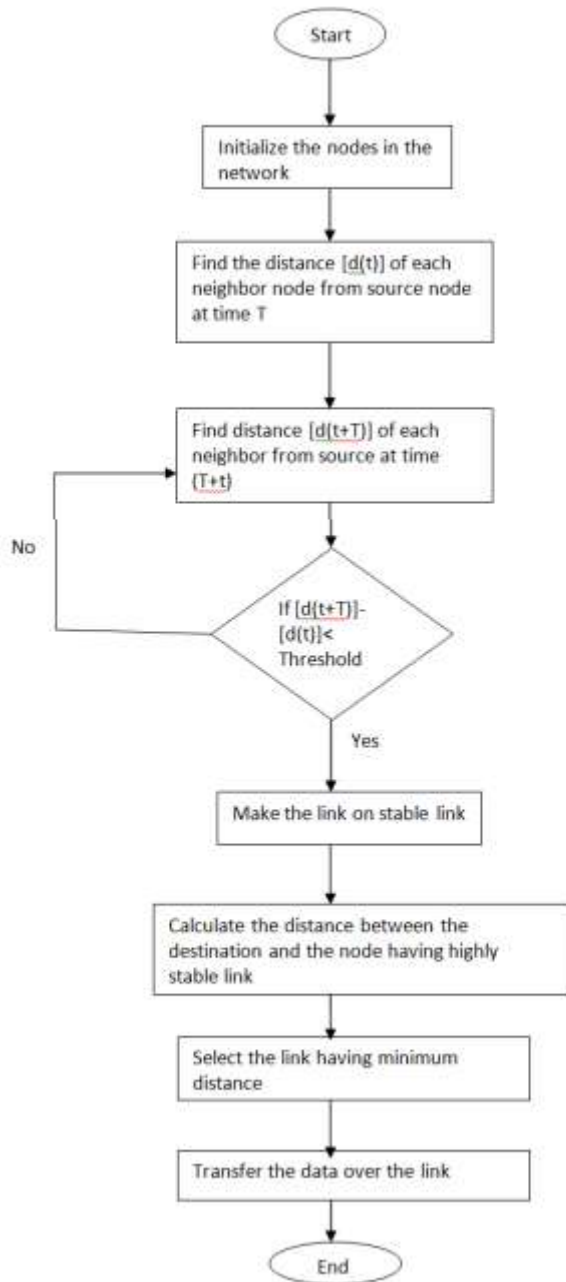


Figure 4. Flowchart for the proposed APU system

The Flowchart (fig.4) represents the methodology of the work done. Initially the nodes in the network have been deployed in such a way that all the nodes are location aware. All location aware nodes broadcast beacons to their neighbor nodes to calculate the distance from source at time T. After that the distance at time (t+T) is calculated. If the difference between the calculated distances is less than threshold then it is considered as stable link. Select the link having the minimum distance by calculating the distance between destination node

and highly stable link and then data can be transferred over the network.

#### 4. RESULTS AND DISCUSSION

Proposed APU system is evaluated on the following parameters:

**4.1 PDR (Packets Delivery Ratio):** PDR is the proportion to the total amount of packets reached the receiver and amount of packet sent by source. If the amount of malicious node increases, PDR decreases. The higher mobility of nodes causes PDR to decrease.

$$PDR (\%) = \frac{\text{Number of packets successfully delivered to destination}}{\text{Number of packets generated by source node}}$$

**4.2 Energy Consumption:** It is the amount of energy consumed by the sensors for the data transmission over the network

Energy Consumption = Sum of energy consumed by each sensor.

**4.3 Overhead:** It is defined as the number of messages involved in beacon update process.

Overhead = No. of messages involved in beacon update process

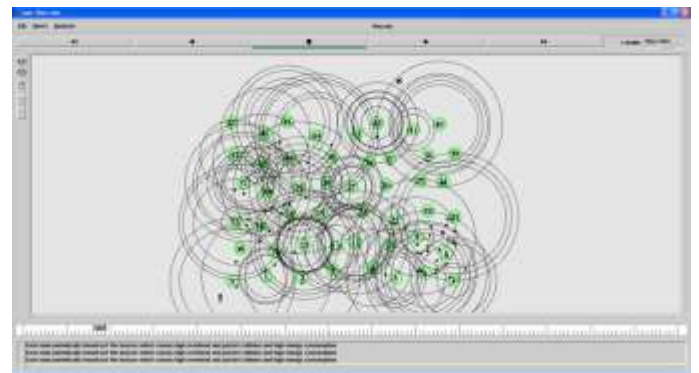


Figure 5. Implementation of the proposed APU system in NS2 environment

Above snapshot (fig.5) shows each node periodically broadcast the beacon which causes high overhead and packet collision and high energy consumption.



Figure 6. Node selection in proposed APU system

In given fig.6 Nodes in blue color are MP nodes which are having the actual location with larger difference from its predicted location. Hence MP nodes update the beacon packet. Deviation threshold is fixed as 60m. If there exists difference between actual locations and predicted of a node is greater than 60m then beacon packet is sent by the node.

Table 1. Energy Consumption versus Time

|    | Existing | Proposed |
|----|----------|----------|
| 0  | 0        | 0        |
| 2  | 75       | 40       |
| 4  | 150      | 75       |
| 6  | 225      | 120      |
| 8  | 310      | 160      |
| 10 | 380      | 190      |
| 12 | 460      | 230      |
| 14 | 540      | 275      |

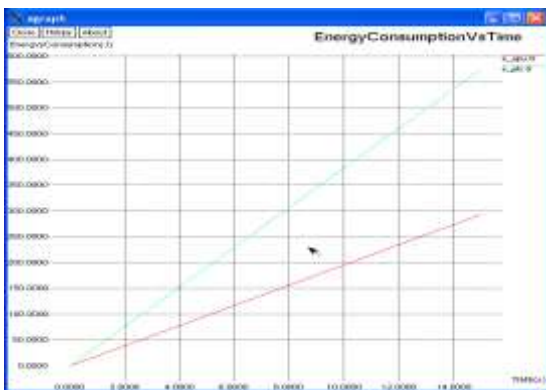


Figure 7. graph showing energy consumption in existing scheme and proposed APU scheme

Given fig.7 shows the energy consumption in existing scheme and proposed APU scheme and the table1 showing the values of graphs.

Energy consumption in existing scheme is high compared to proposed APU since periodic beacon causes high energy consumption in the nodes. APU saves energy by avoiding unnecessary beacon update and do the beacon update adaptively.

Table 2. Beacon Overhead versus Time

|    | Existing | Proposed |
|----|----------|----------|
| 0  | 0        | 0        |
| 2  | 14       | 0        |
| 4  | 27       | 1        |
| 6  | 40       | 1        |
| 8  | 54       | 2        |
| 10 | 68       | 3        |
| 12 | 81       | 3        |
| 14 | 95       | 4        |

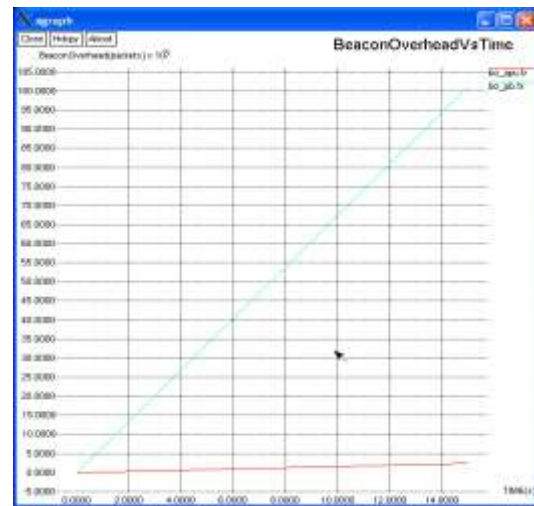


Figure 8. graph showing beacon overhead in existing scheme and proposed APU scheme

Given fig.8 shows the beacon overhead in existing scheme and proposed APU scheme and the table2 showing the values of graphs.

Beacon overhead in existing scheme is high compared to proposed APU due to periodic beacon. APU reduces the beacon overhead by avoiding unnecessary beacon update and only does the beacon update process adaptively.

**Table 3. Packet Delivery Ratio versus Time**

|    | Existing | Proposed |
|----|----------|----------|
| 0  | 0        | 0        |
| 2  | 5        | 12       |
| 4  | 9        | 23       |
| 6  | 14       | 35       |
| 8  | 19       | 47       |
| 10 | 23       | 59       |
| 12 | 27       | 72       |
| 14 | 37       | 83       |



**Figure 9. graph showing beacon overhead in existing scheme and proposed APU scheme**

Given fig.9 shows the packet delivery ratio in existing scheme and proposed APU scheme and the table3 showing the values of graphs.

Packet delivery ratio of proposed APU is high compared to existing scheme. Since network traffic in APU is reduced due to adaptive beacon update instead of periodic beacons in the case of periodic beacon scheme. In PB data gets dropped due to high traffic in the network.

## 5. CONCLUSION AND FUTURE WORK

In this paper, the need to adapt the beacon update is identified and the corresponding policy is employed in geographic routing protocols to the node mobility dynamics and the traffic load. The Adaptive Position Update (APU) strategy is proposed to address these problems. The APU scheme employs two mutually exclusive rules. The MP rule uses mobility prediction to estimate the accuracy of the location estimate and adapts the beacon update interval accordingly, instead of using periodic beaconing. The ODL rule allows nodes along the data forwarding path to maintain an accurate view of the local topology by exchanging beacons in response to data packets that are overheard from new neighbors. Performance of APU is evaluated using extensive NS-2 simulations for varying node speeds and traffic load. Proposed APU strategy leads to cost reduction in beacon update due to destination aware routing and increase in performance over existing system. Results indicate that the APU strategy generates less or similar amount of beacon overhead as other beaconing schemes but achieve better packet delivery ratio, less overhead and energy consumption.

Future work will be the exploring the new techniques to the proposed work to reduce the overhead and energy consumption further in the network.

## 6. REFERENCES

- [1]Sangeeta, Kirti Singh(2011),”Energy Efficient Routing in MANET using OLSR”, International Journal on Computer Science and Engineering, vol. 3, no. 4.
- [2] Smita Singh, Shradha Singh, Soniya jain, S.R. Birdar(2012), “Comparison and study of AOMDV and DSDV Routing Protocols in MANET using NS-2”, International Journal on Computer Science and Engineering, vol. 4, no. 3.
- [3]Anuj K.Gupta, Harsh Sadawarti, Anil K.Verma(2011),“Review of Various Routing Protocols for MANETs”, International Journal of Information and Electronics Engineering, vol. 1, no. 3.
- [4]Gaurav Sachan, D.K.Sharma, Karishma Tyagi, Abhimanyu Parsad(2013), “Enhanced Energy Aware geographic Routing Protocol in MANET: A Review”, International Journal of Modern Engineering Research,vol. 3, pp. 845-848, Issue 2.
- [5]Divya Wadhwa, Deepika, Vimmi Kochher, Rajesh Kumar Tyagi(2014), ”A Review of Comparison of Geographic Routing Protocols in Mobile Adhoc Network”, Advance in Electronic and Electric Engineering, vol.4, pp. 51-58, no.1.
- [6]Varun G Menon, Joe Prathap P M(2013),” Performance Analysis of Geographic Routing Protocol in Highly Mobile Ad hoc Network”, Journal of Theoretical and Applied Information Technology, vol. 54, no. 1.
- [7]Michele Zorzi, Ramesh R. Rao(2003), “Geographic Random Forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance”, *IEEE Transactions on Mobile Computing*, vol. 2, no. 4.

[8]Quan Jun Chen, Salil S. Kanhere, Mahbub Hassan, Kun-Chan Lan(2005), “Adaptive Position Update in Geographic Routing”, School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia.

[9]Dongjin Son, Ahmed Helmy, Bhaskar Krishnamachari,”The Effect of Mobility-induced Location Errors on Geographic Routing in Mobile Ad Hoc and Sensor Networks: Analysis and Improvement using Mobility Prediction”, Department of Electrical Engineering, University of Southern California, 90089.

[10] Pasupuleti Neelima, K.Bhargavi(2014), “Beaconing Strategy for Geographic Routing in Mobile Ad Hoc Network”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 7.

[11]Ravi Kumar Poluru, T. Sunil Kumar Reddy(2014), D.Nagaraju, “Updating the Routing Position in Ad hoc Networks by using Adaptive Position”, International Journal of Computer Trends and Technology (IJCTT), Vol. 10 No. 3.

[12]B. Karp and H. T. Kung., “GPSR: Greedy Perimeter Stateless Routing for Wireless Networks”, in Proceedings of ACM Mobicom 2000, pp. 243-254, Boston, MA, USA,2000.

[13]P. Casari, M. Nati, C. Petrioli and M.Zorzi(2007), “Efficient Non Planar Routing around Dead Ends in Sparse Topologies using Random Forwarding”, in Proceedings of IEEE ICC, pp.3122-3129. Scotland,UK.

[14]Y. Kim, R. Govindan, B. Karp and S. Shenker(2005), “Geographic routing made practical”, in Proceedings of the 2nd Conference on Symposium on Networked Systems Design and Implementation, pp. 217-230, Berkeley, CA, USA.

[15]Quanjun Chen, Salil S. Kanhere, Mahbub Hassan(2013), “Adaptive Position Update for Geographic Routing in Mobile Ad Hoc Networks”, IEEE Transactions on Mobile Computing, vol. 12, no. 3.

[16]J. Hightower and G. Borriello(2001), “Location System for Ubiquitous Computing”, IEEE computer, vol. 34, pp. 57-66, no. 8.

[17]Paolo Casari, Michele Nati, Chiara Petrioli, Michele Zorzi(2007), “Efficient Non-Planar Routing around Dead Ends in Sparse Topologies using Random Forwarding”, IEEE Communications Society, pp. 3122-3129.