



Faculty of Computing, Engineering and Technology

# JOURNAL OF APPLIED TECHNOLOGY AND INNOVATION

Volume 1, Issue 1 July, 2017



## Journal of Applied Technology and Innovation

#### Volume 1, Issue 1 (2017)

Editor-in-Chief Dr. Veeraiyah Thangasamy (APU, Malaysia)

Associate Editors Prof. Dr. Ir. Vinesh Thiruchelvam (APU, Malaysia) Prof. Ir. Dr. Wong Hin-Yong (MMU, Malaysia) Prof. Dr. Quan Min Zhu (UWE, United Kingdom) Assoc. Prof. Ir. Dr. Mandeep Singh (UKM, Malaysia) Dr. Thang Ka Fei (APU, Malaysia) Dr. Lai Nai Shyan (APU, Malaysia) Dr. Babak Basharirad (APU, Malaysia) Dr. Maryam Shahpasand (APU, Malaysia) Dr. Imran Medi (APU, Malaysia) Dr. Noor Ain Kamsani (UPM, Malaysia) Dr. Zubaida Yusoff (MMU, Malaysia) Ir. Dr. Dhakshyani (APU, Malaysia) Mr. Shankar Duraikannan (APU, Malaysia) Assoc. Prof. Dr. Sreeja (SSN, India)

©APU Press, APU 2016

Journal of Applied Technology and Innovation (JATI) is an electronic, open access, peer-reviewed journal that publishes original articles on novel theories, methods and applications in the field of electrical, electronics, mechatronics, telecommunication, computer, information technology, and software engineering. JATI e-journal reviews articles approximately in four (4) weeks period and publishes accepted articles, upon receiving the final versions on the forthcoming issue. It publishes 4 issues per year.

All rights reserved. No part of this publication may be reproduced, copied and stored in any retrieval system or transmitted in any form or by any means, electronically, photocopying, recording or otherwise, without prior permission in writing from the Director of APU Press, Asia Pacific University of Technology & Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia. Email: jati.editor@apu.edu.my

No	Title & Authors	Page
1	A Short Literature Review on the Internet of Things: Research and Development Projects Leticia Nkulu Nsenga Babak Bashari Rad	1-9
2	<b>Essential Functions for Localization in Wireless Sensor Networks</b> <b>Using Geographic Coordinates</b> Thomas O'Daniel	10-27
3	A Review on Social Media Issues and Security Awareness among the users Nor Azlina Abd Rahman Fauziah Permatasari Yuni Hafsari	28-36
4	An Insight into Programming Paradigms and It's Programming Languages M. Selvakumar Samuel	37-57
5	<b>A Review of Data Analytical Approaches in the Insurance Industry</b> Noorhannah Boodhun	58-73
6	<b>Wireless Power Supply for Portable Devices</b> Ahmed Najib Bhutta Veeraiyah Thangasamy Chandrasekharan Nataraj	74-87



# A Short Literature Review on the Internet of Things: Research and Development Projects

Leticia Nkulu Nsenga School of Computing Asia Pacific University of Technology & Innovation (APU) Technology Park Malaysia, Bukit Jalil, Kuala Lumpur 57000 Malaysia leticia.nkulu@gmail.com

Babak Bashari Rad School of Computing Asia Pacific University of Technology & Innovation (APU) Technology Park Malaysia, Bukit Jalil, Kuala Lumpur 57000 Malaysia babak.basharirad@apu.edu.my

**Abstract** - There have been cases, where projects implemented the technological innovation early upon after introduction, while the technology field was still new and knowledge was not matured. The projects were predicted to fail, since the technology was still undeveloped, and that prediction came true. The same downfall has been suspected to happen to the field of IoT as well. This paper carries information on measures that have been taken in the areas of research, innovation and deployment IoT Solutions, an overview of projects funded by the European commission such as the iCORE project, COMPOSE project, the SmartSantader project and the Open IoT project, and how the Thing Commandment principles can be implemented to ensure better deployment of IoT solutions leading to the success of IoT projects.

Index Terms - Internet of Things (IoT); IoT architecture; project development

## 1. Introduction

Information has been an important part of our lives for centuries. This information gradually has been converted to knowledge, and people used that knowledge to improve their lives in various ways. It is the fuel on which the world is running on, today. The existence of the internet has been an important addition to the communication technologies, such as the television and the radio in the sharing of information. With the internet, individuals can choose the type of information they would like to access, and their knowledge on various topics is not subjected to control from the media. The internet can be compared to the circulatory system of the world. From it, information from every topic can be shared and is made accessible for everyone to use. This circulatory system has evolved over the years is no longer connecting people and sharing their information, but "things", which are various forms of computers [1].



These days, computers are taking over jobs that humans have been doing originally; mostly, for manual labour that is done routinely. However, it is said to be possible for some cognitive tasks as well. For example the Enlitic's system that is used to determine if patients have cancer tumours by examining and comparing their CT scans is said to the better at the job than real doctors, with a 0% error rate, while real doctors are said to have an error rate of 7% [2]. According to a study by Benedikt Frey et al. (2013) on the computerization of jobs, it was found out that 47% of American jobs had a chance of being automated using computers. Especially those under transportation, logistics, sales and services as well as support jobs. While in Britain and Japan, this probability was 35% and 49% respectively [3].

Even though job automation and computerization is creating a threat to the job security of people in these countries, the benefits obtained from this is increased productivity, efficiency and accuracy and reduced cost for businesses or organizations [3]. This exceptional job performance has encouraged the connection of these computers to the internet, allowing them to communicate and increasing those benefits even more. The result is a new world where computers and people are all connected and share information. It is what is now called "The internet of things" (IoT).

A research conducted by Gartner; a popular information technology research and advisory company, about 20 years ago, predicted that most organizations that implemented new technology innovation soon after their introduction were going to fail in future due to the immature knowledge about the new technologies. This prediction came true, and many of those projects failed. This trend is suspected to also happen to IoT as well, reported by Sushil Pramanick, an IoT Leader for IBM analytics [4]. IoT is getting implemented rapidly to develop smart cities, industries, and very soon, homes, health and other domains will catch up. The aim of internet of things is to connect the environment, people, and objects. If IoT architectures fail just like some other previous projects, in the future, then it will create a domino effect of problems across all application domains involved that can lead catastrophic disasters. If the knowledge about IoT is improved, then all stakeholders will be able to avoid being affected.

## 2. IoT Architecture

The network architecture of IoT consists of 4 main layers [5]. It consists of the sensing layer also known as the device layer, the network and communication layer and a service support and application support layer. Fig. 1 depicts the architecture of IoT, briefly.

The sensing layer/ device layer is the interface where interaction between the physical world and the whole IoT network architecture is found [5]. The key components of the sensing layer are sensors. The sensor can detect either the presence of light , proximity or position, motion, velocity, and displacement, temperature, sound, pressure or load, gas, chemical substance and electricity [6].

Application layer is a layer that holds software applications that process data collected from sensors. These applications can vary depending on the type of data that is



collected by the network. The software can be fall in the domain of healthcare for example medical monitoring software, precision agriculture, transportation, supply chain management, disaster monitoring [5].

Key technologies that make up the IoT architecture include cloud computing, wireless communication, RFID technology, sensor technology, and advance internet protocol IPV6 [5]. Network layer that provide networking and transport capabilities to the IoT network [7]. Service support and application support layer is the part that deals with IoT service resolution [7].



Fig. 1: IoT Architecture [7]

The internet of things (IoT) is now the solution to many world problems. It is applied in various domains such agriculture, healthcare, home, industries. Most of IoT solutions have been implemented in industrialized or developed nations such Japan, some countries in Europe, the USA. Examples of these IoT solutions include smart homes, smart cities, smart health, smart buildings [6]. Coming up with IoT solutions requires research into problems from an application domain, and research in available technologies to invent a solution. If the solution is deemed successful it is then deployed into a market relating to that domain [7].



## 3. Early Adoption of Innovative Technology

When new technology creates a "hype" once introduced to the public, people would want to use it [8]. In this century, the use of computer technology has become a vital part of many lives. Businesses are the ones that demand technology the most to maintain their competitive edge. This demand urges developers of technology to produce quick solutions to meet that high demand, once the idea of that technology catches their interest. According to the research by Gartner on several projects from organizations that did an early adoption of new information technology innovations, predictions were made that these projects were likely fail after a period of time, because the knowledge behind that technology was not mature enough. Developers and engineers chose to dive into the development of these technology solutions due to the high demand from businesses organizations without conducting enough research to ensure that the solution they are providing, was going to guarantee them a long-term success [9].

According to Gartner, about 20 years ago, there used to be a "hype" about data warehouse and business intelligence [4]. Many organizations were deciding to use decentralized reporting solutions, basic analytic solutions, or a combination of both. Unfortunately, after implementation, lots of problems relating to the quality of data were created, since they had built data warehouse platforms that were disconnected. This led to integration problems as well. Based on findings in 2005, half of these data warehouse projects had failed [9]. More projects continued to fail and by 2012, only 30% were found to exist [4].

According to Sushil Pramanick, companies are trying to embark on Internet of Things initiatives using very narrow, point-focused solutions with very little enterprise IoT strategy in place, and in some cases, engaging or building unproven solution architectures [4].

Since the motion has already been set and many organizations have done an early implementation of IoT technology, it is less likely that they will go back to old technology. No company would want to do that especially if there has been high profit obtained from it and they were leading among other companies in a market. It is also not possible for these organizations would want to abandon their implemented IoT solutions until proven ones have been made available, or until some planned strategy for the IoT enterprises exist.

## 4. Challenges in IoT

There are lots of challenges (shown in Fig. 2) that must be tackled in IoT. Under the technical category, there are challenges relating to reliability, and connectivity. In the policy category, there are issues relating to data, and legacy regulatory models. The third group of issues overlap between the both categories, such as standards, interoperability, and security [6].





Fig. 2: Challenges faced in IoT [6].

When an IoT solutions comes into existence, it means there was a problem that needed solution. To solve any kind of problem, one needs to define it well, list down all possible solutions, and then choose the best of those solutions for implementation. To produce new IoT solutions, all these steps would require research.

Research can be performed at all layers of the IoT architecture. There are problems that require IoT solution in various application domains like healthcare to improve quality of services, environment management and monitoring, for example to ensure sustainability of resources, home and city automation to improve security, comfort, and efficiency in managing the cost of electricity and water, manufacturing industries to reduce the cost of manufacturing processes and to improve their efficiency [10].

Fortunately, in Europe, the European Commission (EC) has been trying to get people from different domains to work together so to promote the growth of joint research projects, as well as new innovations. This is a valuable strategy, because by placing people from different domain, all forms of research questions can be asked, and answered, producing more efficient results that will lead to more reliable innovative IoT solutions. Their biggest concern is to remove the problems caused by fragmentation found across technologies, for example network technologies, cloud computing technologies and application domains such as agriculture, environment, and health care. The EC has created a research and innovation agenda that holds all the all discussions about IoT project they are working on. Trends in technology are document in the agenda, application of those technologies, list of technology enablers, and lists of prioritized research[7].

It also holds the stakeholders taking part in the IoT European Research Cluster (IERC); which is an organization in Europe that tries to direct their attention and effort towards areas that hold great potential for IoT-based capabilities (IERC 2016).



## 5. IoT Project Development

Vermesan and Friess state that projects have four phases. The design phase; which is what happens before starting the project, the executions stage; what happens during the project, the result phase; which is what happens when a project ends, the acceptance and sustainability phase which is what happens after the project ends [7].

Design phase: Main problems or issues will be given special attention to, depending on what the purpose of the project is.

Execution phase: During the project, plans get to change based on the type problems that are encountered, or based on rising opportunities.

Result phase: A summary of what had been achieved is made, then they will be compared against expected results of the project. Special attention will be given to any lessons learned.

The acceptance and sustainability phase: This is the stage that is important to ensure success of the IoT solution into the market. There will be a sustainability plan prepared for problems that arise due to involvement of various stakeholders in IoT. Depending on what type of domain the IoT solution was designed for, this is will to help prepare for economic, social, or production process related problems.

## 6. Review of EC projects

The EC has funded many project for the advancement of IoT, but 3 projects have been highlighted to be most successful. These include the COMPOSE project, Smart Santander, and Open IoT [7].

## 6.1 The iCORE Project

It was a project that was enable the use of more cognitive technologies in IoT and making IoT based applications responsive to the need of users and adaptable to changes of those needs.

#### 6.2 COMPOSE

The aim of this project was to enable integration between IoT and the Internet of Services (IoS) through a platform, where data from objects connected to the internet can be published, shared and used by other applications and services. It is a great tool that can be used by developers interested in developing smart applications that can communicate with smart "things" or devices, and accessing information from external sources.



#### 6.3 SmartSantander

This project was created for the purpose of testing architectures under research or being experimented, testing important enabling technologies, as well as services to be used in IoT. This project had made Europe a leader in Internet of things technologies, as it enabled people in the science community to perform their experiments and do evaluations of services and applications for smart cities under real-life conditions.

#### 6.4 Open IOT

This is a project that had been developed to encourage open source contributors and IoT researchers to work together. There are lots of community building activities. This project has been used to provide training services, support services.

Some researchers have tried to collect information on what it takes to move from research and innovation to market deployment using the above-mentioned projects [7]. However, there has been detailed information and emphasis on the research and innovation than deployment. Perhaps it is due to the extensive technical detail that would have to fill the documentations.

## 7. The Thing Commandments

Even though these can be considered new, a set of principles that can be applied to the internet of things had been proposed by Edewede Oriwoh, Paul Sant, and Gregory Epiphaniou from the University of Bedfordshire, UK. The aim of these principles is to give a guideline from which people can refer to, to make the right decisions and making them aware of the requirements for making the right choices, however these principles are focused on security. The principles can be applied by manufacturers and developers, and individual consumers. These principles are known as the Thing commandments [1].

Principle 1: Your thing is your thing

This principle states that objects belong to their owners, and owners should be responsible for them. In case, that object performs something wrong, the owner should be held responsible. This is to ensure that in a case a crime is committed using that object, the owner should bear some form of responsibility.

Principle 2: Your thing should have a relationship with all other things

All things of a particular owner should fall under one network, and they should be able to identify one another. Application of this principle will ensure that access is denied to any other objects that do not belong to the owner.



Principle 3: Your thing should be able to identify communication between each other by the use of an established method

The purpose of this principle is to make sure that no foreign objects are permitted access to the owner's network, and they should be detected as they try to communicate with the network.

Principle 4: Usability and ease of use

Owners should be able to perform configurations seamlessly, and make modifications such as adding new objects or "things" and removing them when they wish to. There should be no need for training, or service requirements from specialist vendors.

Principle 5: All things should be controllable by their owner

Vendors should apply this principle. They should ensure that all technologies used by the owners can be controlled fully by them. Those technologies should be user-friendly and should be easy to understand.

Principle 6: Everything should have an owner

Everything should hold a form of identification pertaining to the owner. The purpose of this is to help law enforcement agencies as they try to perform investigations in case a crime occurs.

Principle 7: Refusing, disabling and destroying or disposing of things

A person should not be forced to use objects that they do not approve of, and that it should not be difficult for people to dispose their objects when they need to. Once the object is destroyed, all data contained in that object should be destroyed as well.

## 8. Conclusion

Statistical research has shown the negative impact that early adoption of innovative technology had on previous information technology projects 20 years ago. Organizations such as the EC are working towards funding research and innovation projects to ensure the field of IoT technology stays successful. This shows awareness of the past failing technologies and importance of what role IoT will play in future for business and individuals. Various challenges faced in IoT are being given attention to by the European commissions. They do have a great involvement in the advancement of IoT such as encouraging stakeholders from the IERC in taking part in the endeavour, as well taking part in research. This has led to the existence of Innovative projects such as the iCORE project, COMPOSE project, the SmartSantader project and the Open IoT project.

Moreover, many scholars have also placed some efforts in contributing towards the success of IoT. Contributions from various sources will increase the chances of a brighter future in field of IoT. A list of principles that are worth of implementation for improved deployment have been proposed.



## References

- [1] E. Oriwoh, P. Sant, and G. Epiphaniou, "Guidelines for internet of things deployment approaches--the thing commandments," *Procedia Computer Science*, vol. 21, pp. 122-131, 2013.
- [2] Economist, "Artificial intelligence: The impact on jobs Automation and anxiety," in *The Economist*, ed. London The Economist Newspaper Limited, 2016.
- [3] C. B. Frey and M. A. Osborne, "The future of employment: how susceptible are jobs to computerisation?," *Technological Forecasting and Social Change*, vol. 114, pp. 254-280, 2017.
- [4] S. Pramanick, "Why over two-thirds of the Internet of Things projects will fail," in *Big Data & Analytics Hub*, ed: IBM, 2016.
- [5] X.-Y. Chen and Z.-G. Jin, "Research on key technology and applications for internet of things," *Physics Procedia*, vol. 33, pp. 561-566, 2012.
- [6] P. Biggs, J. Garrity, C. LaSalle, and A. Polomska, "Harnessing the Internet of Things for Global Development," *ITU and CISCO*, 2016.
- [7] O. Vermesan and P. Friess, Internet of Things Applications-From Research and Innovation to Market Deployment: The River Publishers, 2014.
- [8] A. McEwen and H. Cassimally, *Designing the internet of things*: John Wiley & Sons, 2013.
- [9] Gartner. (2005, Gartner Says More Than 50 Percent of Data Warehouse Projects Will Have Limited Acceptance or Will Be Failures Through 2007. *Press Release*. Available: <u>http://www.gartner.com/newsroom/id/492112</u>
- [10] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, pp. 1497-1516, 2012.



# Essential Functions for Localization in Wireless Sensor Networks Using Geographic Coordinates

Thomas O'Daniel Faculty of Computing, Engineering, & Technology Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia dr.thomas.odaniel@apu.edu.my

*Abstract*- A variety of localization protocols have been proposed in the literature which allow Wireless Sensor Network (WSN) nodes to interpolate their location from their neighbors as an alternative to deploying more expensive WSN nodes with GPS receivers or other dedicated localization hardware. This paper presents a set of efficient functions applied to three base cases where a WSN node calculates an initial estimate of its location and a finite set of alternate points that could be its actual location, given the GPS coordinates and nominal transmission radius of two or three neighbors. The process of narrowing the set of possible actual locations through iterative refinement as more nodes join the network is discussed, along with the limits on the accuracy of the overall network map.

*Index Terms*- Wireless sensor network (WSN), global positioning system (GPS); localization protocols

## 1. Introduction

Wireless Sensor Networks (WSNs) are a fundamental aspect of ubiquitous systems and the Internet of Things (IoT). WSNs are composed of tiny devices with constrained processing and memory resources that are typically battery powered. Networks of these devices are characterized by small packet payload size, minimum bandwidth, unreliable radio connectivity, ad hoc deployment, dynamic topology changes, and nodes running in a power conservation mode to prolong battery lifetime.

Many industrial applications consist of a large number of randomly distributed nodes, so it is advantageous if the network is able to autonomously build the communication links and control the communication between nodes [1]. WSN deployments for environmental surveillance and disaster management in particular could benefit from constant reporting of the location where data was sensed. Nodes can be equipped with Global Positioning System (GPS), but this is a costly solution in terms



of both money and energy consumption [2] [3], and GPS typically fails inside buildings and under heavy vegetative cover [4].

This paper describes the fundamental calculations necessary for a node to estimate its position given the GPS coordinates of some neighbors and an indication of their transmission radius. The most basic principle of triangulation is that given two points and the distance between them, a third point can be found. Ancient texts record the use of triangulation to estimate distances. Two common examples: to measure the distance from shore to a remote ship, mark two points on the shore with a known distance between them and calculate the angles between this baseline and the location of the ship; to measure the height of a mountain or lighthouse, use the distance between two ground points and the angles to the top.

In surveying *trilateration* is the process of determining absolute or relative locations of points purely by measurement of distances, while the term *triangulation* is reserved for the process that involves only angle measurements, The use of both angle and distance measurements is referred to as *triangulateration* by those who find these distinctions meaningful. *Multilateration* is a technique based on measuring power levels and antenna patterns, commonly used with radio navigation systems. Unlike measurements of absolute distance or angle, using a radio signal to measure the distance between two stations at known locations emitting broadcast signals at known times results in an infinite number of locations that satisfy the "time difference of arrival" metric. Multilateration requires at least three synchronized emitters for determining location in two dimensions, and at least four for three dimensions.

Many WSN localization techniques reported in the literature use various combinations of metrics to develop measures of link quality, but inferring relative location from these measures is subject to assumptions about decrease in signal strength due to the distance between transmitter and receiver, type and height of antennas, and the presence of obstacles that disrupt the line-of-sight path [1] [5] [6]. The techniques presented here simply require each node to have the ability to transmit its actual or presumed location, and its nominal transmission radius. Exactly how this is achieved (through beaconing, addressing, or some type of protocol for example) is not important for the calculations. The calculations are done with locations expressed as decimal GPS coordinates and distances in kilometers; other coordinate systems and distance measurements could be used.

The first section of this paper reviews the basic terms and concepts related to solving triangles and geolocation. The second section presents the essential formulae expressed as functions in the C programming language, which can be easily ported to another. The third section shows how the essential functions can be used by a WSN node to establish an initial estimate of its location given minimal information, along with a finite set of alternate points that could be its actual location. This is followed by an examination of the process for refining the initial estimate using several of the sets of alternate points, and consideration of the limits on overall accuracy.



## 2. Materials and Methods

#### 2.1 Basic Principles of Triangulation

#### 2.1.1 Characteristics of Triangles

Triangles have several interesting properties:

The shortest side is always opposite the smallest interior angle

- $\circ$   $\;$  The longest side is always opposite the largest interior angle  $\;$
- The interior angles of a triangle always add up to 180°
- $\circ~$  The exterior angles of a triangle always add up to 360° thus given three points, it is possible to draw a circle that passes through all three (the circumcircle of the triangle)
- Any side of a triangle is always shorter than the sum of the other two sides; in other words, a triangle cannot be constructed from three line segments if any of them is longer than the sum of the other two. This is known as the Triangle Inequality Theorem.

Solving a triangle means finding the unknown lengths and/or angles. The classic problem is to specify three of the six characteristics (3 sides, 3 angles) and determine the other three. Any combination except 3 angles allows determination of the other side lengths and angles - three angles alone determines the shape of the triangle, but not the size. The actual solution depends on the specific problem, but the same tools are always used:

- $\circ$   $\;$  The knowledge that the sum of the angles of a triangle is 180°.
- The Pythagorean theorem, the essence of which is that for any triangle a line can be drawn that divides it into two right triangles, and the relationship between the sides of a right triangle is such that the square of the length of the longest side equals the sum of the squares of the lengths of the other two sides  $(c^2 = a^2 + b^2)$  in the notation explained below).
- $\circ~$  The trigonometric functions that relate a given angle measure to a given side length.

#### Essential Terminology

It is usual to name each vertex (angle) of a triangle with a single upper-case letter, and name the sides with the lower-case letter corresponding to the opposite angle, as illustrated in Fig. 1. Alternatively, the sides of a triangle can be labeled for the vertices they join, so side b would be called line segment AC.

the vertices they on which side is **Fig. 1** 

The height or altitude of a triangle depends upon which side is selected as the base. An altitude of a triangle is a line through a vertex

of a triangle that meets the opposite side at right angles. This point will be inside the triangle when the longest side is the base; if one of the angles opposite the chosen





vertex is obtuse (greater than 90°), then this point will lie outside the triangle. The area of a triangle is one-half the product of its base and its perpendicular height; in the case of a right triangle, this is the product of the sides that form the right angle.

A special set of terms is used to describe right triangles: the hypotenuse is the longest side, an "opposite" side is the one across from a given angle, and an "adjacent" side is next to a given angle. There are six trigonometric functions that take an angle argument and return the ratio of two of the sides of a right triangle that contain that angle. For any given angle L

- 1. Sine: sin(L) = Opposite / Hypotenuse
- 2. Cosine: cos(L) = Adjacent / Hypotenuse
- 3. Tangent: tan(L) = Opposite / Adjacent

atan(), asin(), and acos() are the respective inverses of tan(), sin(), and cos().

In C, C++, Java, python, and other programming languages the trigonometric functions take a parameter and return a value expressed in radians. The radian is the standard unit of angular measure, used in many areas of mathematics. One radian is the angle at the center of a circle where the arc is equal in length to the radius, as illustrated in Fig. 2 (a). More generally, the magnitude in radians of an angle is the arc length divided by the radius of the circle. As the ratio of two lengths, the radian is a "pure number" that needs no unit symbol.

The number pi is a mathematical constant, the circumference divided by the diameter of any circle. One radian is equivalent to 180 / pi (57.29578) degrees; Fig. 2 (b) illustrates this relationship.

The trigonometric functions actually work with a "unit circle" centered at (0,0) with a radius of one unit, so it intersects the X and Y axes at (1,0), (0,1), (-1,0), and (0,-1). They return a value between 1 and -1, and multiplying this number



by the length of the vector yields the exact Cartesian coordinates of the vector. *Solving Triangles* 

As noted above, solving a triangle means finding the unknown lengths and/or angles. Given any three of the six parameters (except 3 angles without a side length), any triangle can be solved using three equations:

- 4.  $A + B + C = 180^{\circ}$  [Angles sum to 180]
- 5.  $c^2 = a^2 + b^2 2^*a^*b^*cos(C)$  [The Law of Cosines]
- 6. a / sin(A) = b / sin(B) = c / sin(C) [The Law of Sines]

Points worthy of mention are (a) the Law of Cosines reduces to the Pythagorean Theorem in the case of right triangles, and (b) determination of an angle or side directly



from its sine will lead to ambiguities since sin(x) = sin(pi - x), while determination from cosine or tangent will be unambiguous. Many formulae have been derived to avoid the sine ambiguity, but the simplest is to use the half angle formula which yields an unambiguous positive or negative result (by symmetry there are similar expressions for angles B and C).

7.  $\sin(A / 2) = \sqrt{(1 - \cos(A)) / 2)}$ 

For geolocation, plane triangles are adequate under certain circumstances (explained below) but the general case involves solving "spherical triangles". A spherical triangle is fully determined by three of its six characteristics (3 sides and 3 angles), and the basic relations used to solve a problem are similar to those above. However, the key differences are that the sides of a spherical triangle are measured in angular units (radians) rather than linear units, and the sum of the interior angles of a spherical triangle is greater than 180°.

Fig. 3 (a) shows how the intersection of three planes through a sphere forms two spherical triangles, one from the solid lines (foreground) and one from the dotted lines (background). The triangle degenerates into three points with the sum of the angles equal to 3\*pi and the sum of the sides equal to 2\*pi on the unit sphere. Euclid (300BC) Book 11, Proposition 21 provides a rigorous proof, with a corollary that the sum of the angles of a spherical triangle is greater than pi [9,



pp.184]. The amount by which the sum of the three angles exceeds pi is referred to as the "spherical excess".

Labeling points and angles on a spherical triangle follows the normal conventions, as shown in Fig. 3 (b). The basic relations used to solve a spherical triangle are similar to those for a planar triangle: modifications to account for the curvature of the sides and the spherical excess lead to analogous formulae for side lengths and area, a Spherical Law of Cosines, and a "Spherical Pythagorean Theorem" (amongst Napier's Rules). Relevant examples are provided in section III.

#### GPS and Geolocation

The Earth is only approximately spherical, so no single value serves as its natural radius. However, the Earth deviates from a perfect sphere by only a third of a percent, making the sphere model adequate in many contexts. Using the polar minimum of 6,357.75 km and the equatorial maximum of 6,378.14 km, several different ways of modeling the Earth as a sphere yield a mean radius of 6,371 km [10].

GPS coordinates are based on dividing this perfect sphere of the world into 360 degrees of horizontal longitude and 180 degrees of vertical latitude. Each degree of latitude and longitude is divided into sixty minutes, and each minute is divided into



sixty seconds, with fractions of a second offering finer-grained specification of a location. There are 3 common and equivalent formats for expressing location, ddd°mm'ss.ss", ddd°mm.mmm', and ddd.ddddd°, where d, m, and s stand for degrees, minutes, and seconds.

Degrees are expressed as a number between -180 and +180 for longitude, and a number between -90 and +90 for latitude. Zero degrees longitude is an arbitrary line, locations to the west of which are negative, and locations to the east are positive. Zero degrees latitude is the equator, with locations to the north as a positive number, and to the south as a negative number.

On the sphere of the world the longitude lines, also known as meridians, are the same distance apart at the equator and converge at the poles. The meter was originally defined such that ten million of them would span the distance from the equator to a pole, so at the equator each degree of both latitude and longitude represents approximately 111.32 km. Because the meridians get closer together moving from the equator toward either pole, one degree of longitude is multiplied by the cosine of the latitude, decreasing the indicative physical distance as illustrated in Table 1 for coordinates expressed as decimal degrees.

Deci mals	0	1	2	3	4	5	6	7		
Equat or	111 km	11 km	1 km	111 m	11 m	1 m	11 cm	1 cm		
Quito, Ecuador; Maqcapa, Brazil; Kampala, Uganda; Thinadhoo, Maldives; Pontianak, Borneo										
23 N	102. 5	10.25	1 km	102.5	10.25	1 m	10.25	1 cm		
// S	km	km		m	m		cm			
Havana, Cuba; Muscat, Oman; Shantou, China // Sao Paulo, Brazil; Windhoek, Namibia; Alice Springs, Australia										
45 N	78.7	7. 9 km	787 m	78.7 m	7.9 m	78.7	7. 9 cm	7. 9		
// S	km					cm		mm		
Portland OR USA; Limoges, France; Harbin, China // Rio Mayo, Argentina; Dunedin, New Zealand										
67 N	43. 5	4.35	435 m	43.5 m	4.35 m	43.5	4.35	4.35		
// S	km	km				cm	cm	mm		
Coldfoot, AK USA; Repulse Bay, Canada; Inari Finland; Tomtor, Siberia // Adelaide, Casey Station, Antarctica										

 Table 1: Precision of GPS decimal places and indicative locations at particular

 latitudes [11]

It is worth noting that the fourth decimal place is comparable to the typical accuracy of an uncorrected GPS unit with no interference, while accuracy to the fifth



decimal place requires differential correction with commercial GPS units. The seventh decimal place is near the limit of what GPS-based techniques can achieve with painstaking measures [12].

#### 2.2 Essential Functions

To calculate the relative location of a point on the globe, we consider a spherical triangle given point A as longitude xA, latitude yA and point B as longitude xB, latitude yB, and derived point xC, yC. The distance between points A and B is easily calculated from their coordinates, so we need either the distance or the angles between point C and these points to determine its coordinates. Since we know the radius of the earth sphere, the characteristics of the triangle can be expressed as radians and the calculations done on the unit sphere.

The essential formulae are below as functions in the C programming language, which can be easily ported to another. One "peculiarity" of C is its lack of a built-in operator for exponentiation, because exponentiation not a primitive operation for most CPUs. Thus a function or a preprocessor macro such as **#define** SQ(v) ((v)\*(v)) is necessary to improve clarity. It is also convenient to have a static constant or preprocessor macro such as **#define** D2R 0.017453293 for converting decimal degrees to radians (this is a library function in Java and python). Similarly, useful constants are **#define** K2R 0.00015696 for converting kilometers to radians and **#define** R\_KM 6371 for the earth radius.

Indispensable references for the spherical earth formulae are Williams [13] where they are presented in a manner that facilitates practical calculation, and Osborn [9] which has the full proofs. For those who are interested, Veness [14] has implemented them in Javascript, along with additional calculations based on an elliptical earth model.

#### 2.2.1 Distance Between Points

The planar linear distance between points A and B given their longitude (x) and latitude (y) is calculated as

```
double lenplnr(double xA, double yA, double xB, double yB) {
```

return sqrt( (SQ(xB - xA)) + (SQ(yB - yA)) ); }

The Haversine formula returns the geodistance between the points in radians, accurate to around 0.3% because it is based on a spherical earth model. It is preferred to the spherical law of cosines because it maintains its accuracy at very small earth distances.

double lenhsine(double xA, double yA, double xB, double yB) {

```
double sinlon = sin( ( (xB - xA) * D2R )/2 );
double sinlat = sin( ( (yB - yA) * D2R )/2 );
return 2 * asin(sqrt(
```



```
(SQ(sinlon)) + cos(xB*D2R) * cos(xA*D2R) * (SQ(sinlat)) )); }
```

#### 2.2.2 Area of a Triangle Given Side Lengths

```
Heron's (aka Hero's) formula is used for the planar triangle
```

```
double areapt(double lenA, double lenB, double lenC) {
```

double s = (a + b + c) / 2;

return sqrt( s \* (s-a) \* (s-b) \* (s-c) ); }

l'Huiller's formula for a spherical triangle is analogous to Heron's for a plane triangle, and maintains its accuracy with small triangles. Argument and return values are in radians, multiply the returned value by SQ(R\_KM) for the surface area enclosed by the triangle.

```
double areast(double lenA, double lenB, double lenC) {
   double s = (lenA + lenB + lenC) / 2;
   return 4 * atan(sqrt(
      tan(s/2) * tan((s-lenA)/2) * tan((s-lenB)/2) * tan((s-lenC)/2)) )); }
```

#### 2.2.3 Side Length of a Right Triangle

For the planar right triangle in Fig. 1 above, given the length of sides a and c the Pythagorean Theorem yields the length of the hypotenuse (side b) as

double lenrpthyp(double lenA, double lenC) {

```
return sqrt( (SQ(lenA)) + (SQ(lenC)) ); }
```

Alternatively, given the length of the hypotenuse and one other side, the length of the third side is

double lenrptside(double lenB, double lenC) {

```
return sqrt( fabs( (SQ(lenB)) - (SQ(lenC)) ) ); }
```

For a spherical triangle with one right angle, there are ten relations (Napier's rules) that allow computing any unknown side or angle in terms of any two of the others. One of these uses the lengths of the sides that form the right angle: a Spherical Pythagorean Theorem.

double lenrsthyp(double lenA, double lenC) {

```
return cos(lenA) * cos(lenC); }
```

or given the length of the hypotenuse and one other side, the length of the third side is double lenrstside(double lenB, double lenC) {

```
return cos(lenB) / cos(lenC); }
```



#### 2.2.4 The Special Case of Small Spherical Triangles

Spherical triangles with side lengths much less than the radius have a spherical excess so small they may be treated as planar. Legendre's theorem shows the angles of the spherical triangle exceed the corresponding angles of the planar triangle by approximately one third of the spherical excess when the side lengths of the spherical triangle are much less than 1 radian. For those who want to check, an intermediate point in the proof of Legendre's theorem presented in [9, equation D48, pp.201] is calculation of the area of the counterpart triangle using the side lengths.

/\* Legendre's theorem - convert the area of a spherical triangle

to the area of a planar triangle with sides of the same length \*/

double areaptst(double starea, double lenA, double lenB, double lenC) {

return (starea / (1 + (( (SQ(lenA)) + (SQ(lenB)) + (SQ(lenC)) )/24))); }

/\* Legendre's theorem - planar to spherical \*/

double areastpt(double ptarea, double lenA, double lenB, double lenC) {

```
return (ptarea * (1 + (( (SQ(lenA)) + (SQ(lenB)) + (SQ(lenC)) )/24))); }
```

#### 2.2.5 Additional Formulae: Intersection of Circles

This calculation [15] saves a lot of work for these scenarios relative to using the triangle formulae above, which could be used to get the same result. Arguments are the radius of the two circles and the distance between their center points, the coordinates of the center points, and two-point (x,y) data structures passed by reference. The function effectively returns the two points where the circles intersect. It is presumed that the length of line PQ is less than the sum of the radius of the circles, so they actually do intersect (recall the Triangle Inequality Theorem).



```
/* offsets of the intersection points from (xH,yH) */
double xVN = -difyPQ * (lenHN/lenPQ);
double yVN = difxPQ * (lenHN/lenPQ);
/* the actual intersection points */
nxy->xcoord = xH + xVN;
nxy->ycoord = yH + yVN;
vxy->xcoord = xH - xVN;
vxy->ycoord = yH - yVN; }
```

#### 2.2.6 Additional Formulae: Points on a Line

These useful functions take an argument of a point (x,y) data structure passed by reference; they could just as easily return this data structure.

```
/* point J on line I--J--K using lenIJ */
void ptada(double xI, double yI, double xK, double yK,
      double lenIJ, double lenIK, struct RETpoint* retxy) {
 retxy->xcoord = xI + ( (lenIJ / lenIK) * (xK - xI) );
 retxy->ycoord = yI + ( (lenIJ / lenIK) * (yK - yI) ); }
/* point J on line I--J--K using lenKJ */
void ptaba(double xI, double yI, double xK, double yK,
      double lenII, double lenIK, struct RETpoint* retxy) {
 retxy->xcoord = xI + ( (lenI] / lenIK) * (xI - xK) );
 retxy->ycoord = yI + ( (lenIJ / lenIK) * (yI - yK) ); }
/* point K on line I--J--K using lenIK */
void ptbdab(double xI, double yI, double xJ, double yJ,
 double lenIK, struct RETpoint* retxy) {
 double ikx = xJ - xI;
 double iky = yJ - yI;
 double bb = sqrt( (SQ(lenIK)) / ((SQ(ikx)) + (SQ(iky))));
 retxy->xcoord = xI + (ikx * bb);
 retxy->ycoord = yI + (iky * bb); }
```

## 3. WSN Node Geolocation

On the Earth the excess of an equilateral spherical triangle with sides 21.3km (and area 393km2) is approximately 1 arc second (1/3600th of a degree). Taking account of both the convergence of the meridians and the curvature of the parallels, if the distance between points is around 20 km the planar distance formula will result in a maximum



error of 30 meters (0.0015%) at 70 degrees latitude, 20 meters at 50 degrees latitude, 9 meters at 30 degrees latitude, and be precise at the equator [16]. From another perspective, at a height of two meters the clear line of sight is around 5 km due to the curvature of the earth, so the planar and spherical calculations would return the same result at any latitude. These are microdistances relative to the earth radius, so the choice of using spherical or planar triangle calculations is open, as long as the absolute necessity of using radians for functions that require then is kept firmly in mind.

Like the spherical earth model, an acceptable simplification for the initial calculations is to show the transmission radius of the wireless sensor network node as a circle. In actuality the transmission radius is irregular as it is subject to various types of interference and dependent on antenna characteristics, but these variables can be left for refinement suitable to specific deployments.

The scenarios presented here are base cases, working with minimal information. The goal is for a node to establish an initial estimate of its location, with a finite set of alternate points that could be its actual location. As more nodes join the network and go through this process more information becomes available, and the nodes can narrow their set of possible actual locations through a process of iterative refinement (discussed below). Ultimately the nodes in the network will be able to converge on a stable network map within a quantifiable margin of error for each node.

The base cases take advantage of the fact that wireless networks are inherently broadcast networks, so every node within range of a given node can hear all transmissions. This leads to the concept of "audible" and "inaudible" neighbors: nodes that can send to and receive from each other are audible neighbors, while a node that can hear its neighbor send to another node but cannot hear the response (i.e., eavesdrop on one side of the conversation) has an inaudible neighbor.

The base cases are also predicated upon the ability of a node to transmit its actual or presumed location, and its nominal transmission radius. Optimally the method will provide a way for a node to communicate the location and transmission radius of its audible and inaudible neighbors as well. Exactly how this is achieved (through beaconing, addressing, or some type of protocol for example) is not important for the calculations. The calculations are done with locations expressed as decimal GPS coordinates and distances in kilometers; other coordinate systems and distance measurements could be used.

All of the radios in the scenarios have an equal transmission radius, to avoid a situation where a radio has a transmission radius that can be completely contained within another – this situation has too high a degree of ambiguity to consider here. The diagrams are all drawn in a manner that would make it easy to superimpose a xy axis for easier comprehension; two-dimensional rotation would only affect the absolute values of the coordinates.



#### 3.1 Base Case: Two Audible Neighbors

In this scenario, radio "dot" can communicate with (send to and receive from) radios P and Q, but P and Q cannot communicate with each other. In other words, P and Q are audible neighbors of "dot" and "dot" is their audible neighbor, while P and Q are inaudible neighbors of each other. Figure 4 shows three variations.

Radio "dot" can interpolate its location from the location coordinates of P and Q and their transmission radius. If "dot" positions itself at an intersection point of the two circles (N or M), it could not move farther away without losing contact but it could move closer, within the area of intersection of the two circles.

Points N and M are returned by the circpts() function, which in fact calculates these points using the height of a triangle with a base side length equal to the distance between P and Q, and the other two sides equal to their transmission radius. The area defined by the spherical triangle NEF or MEF defines the set of possible alternative locations for radio "dot". However, without more information, "dot" cannot know which of N or M it should choose as its location. Nonetheless, a finite set of possibilities has been defined and an arbitrary choice between N and M (Fig. 4 (a) or (b)) must be made until further information is available.

The set of possible locations is inversely proportional to the distance between P and Q: the shorter the distance between them, the greater the area of the triangle becomes. As illustrated in Fig. 4 (c), the calculations are the same when P and Q are audible neighbors, the set of alternative locations just gets larger.

#### 3.2 Base Case: One Audible Neighbor with Two Neighbors

This case is built on the previous one, after radio "dot" has arbitrarily chosen its

position as N. In the first variant, Fig. 5 (a), radio N is the audible neighbor of the new radio "dot", which positions itself at point R; in the other variant (Fig. 5 (b)) radio Q is the audible neighbor of the new radio "dot", which positions itself at point S.

The new radio "dot" uses the intersections of the inaudible neighbor









circles (P and Q in the first case, P and N in the second) to obtain two points, and extends the line from one of these points through the coordinates to a point that is its transmission radius away from the audible neighbor.

The variants are only distinguished by the location of the point returned by the circpts() function that is closest to the audible neighbor: as illustrated in Fig. 5 (a), in the NR variant this point is exactly N, in the QS variant (Fig. 5 (b)) it is not quite exactly Q. Thus it is important to recognize that for this scenario only one of the points returned is useful – the point farther away from the audible neighbor.

In both variants the set of possible alternative points is calculated in the same manner: calculate the overlap of the circle of the audible neighbor with each inaudible neighbor (PN and QN in the first case, QP and QN in the second) and use the points that are farthest apart from each other. In the first case this yields a set of possible actual locations for R as the sum of the areas of triangles RNE and RNF as shown in Fig. 6



(a), in the second it is the sum of the areas of SQE and SQF as shown in Fig. 6 (b). The area is relatively large, but finite for all practical purposes.

#### 3.3 Base Case: One Audible Neighbor with One Audible Neighbor

This case, illustrated in Fig. 7, is likely to arise for edge nodes in particular. The new radio "dot" positions itself at point S by simply extending line PQ by its transmission radius.

In this case triangle QEF defines the inverse of the set of points for the alternative locations: the actual



location of S is any point at distance less than or equal to the transmission radius of S from Q, and outside triangle EFQ.

## 4. Refining the Estimate

Using the triangles that define the set of alternative locations to refine the initial location estimate is where this exercise gets interesting. In principle, the labeled points in Fig. 8 represent five iterations of the initial calculations using different pairs of audible neighbors, but the diagram is made to illustrate some key ideas rather than represent the outcome of a realistic application of the base case.



It is essential to keep in mind that the proximity of the calculated points offers no insight: any point inside the associated triangle is an equally valid (and equally probable) location for the node, since the calculated points simply represent limits on how far away the nodes can be from each other. Refining the estimate involves examining the overlapping areas of the triangles; ideally they would all overlap and yield a very small set of possible alternative locations, but an outcome like the



```
Fig. 8
```

one illustrated in Fig. 8 where no single point satisfies all of the constraints is theoretically possible.

In any case, creating this mapping requires choosing a pair of triangles and either (a) checking to see if any of the points of one lie inside the area of the other, or (b) checking if the sides of one intersect sides of the other. The essential calculations are quite similar, and rely on checking the sign of the vector cross product. Put very simply, the cross product of two vectors is another vector that is at right angles to both.

/\* Vector Cross Product \*/

double vcp(double xP, double yP,

double xF, double yF, double xG, double yG) {

return ((xP - xG)\*(yF - yG) - (xF - xG)\*(yP - yG)); }

/\* point P is inside a triangle if it is on the same side of each line \*/

int ispint(double xP,double yP, double xA,double yA,

double xB,double yB, double xC,double yC) {

```
int vs = 0;
```

```
vs += ( vcp(xP,yP, xA,yA, xB,yB) < 0 ? 1:0 );
```

```
vs += ( vcp(xP,yP, xB,yB, xC,yC) < 0 ? 1:0 );
```

```
vs += ( vcp(xP,yP, xC,yC, xA,yA) < 0 ? 1:0 );
```

```
return ( ((vs == 3) || (vs == 0)) ? 1:0 ); }
```

/\* segment PQ intersects AB when P and Q are on opposite sides of AB

bonus: if not, which side is PQ on, or is this effectively zero \*/

int segint(double xP,double yP, double xQ,double yQ,

```
double xA,double yA, double xB,double yB) {
```

```
double p = vcp(xP,yP, xA,yA, xB,yB);
```

double q = vcp(xQ,yQ, xA,yA, xB,yB);

if  $((p > 0) \&\& (q > 0)) {$ 

if ( (EZERO(p)) && (EZERO(q)) ) return 2;



```
else return 1;
}
if ( (p < 0) && (q < 0) ) {
    if ((EZERO(-p)) && (EZERO(-q))) return -2;
    else return -1;
    }
return 0; }</pre>
```

The segint() function uses what is commonly known as an "epsilon test" rather than testing if the value is zero because of the accuracy of the calculations. Calculations with a variable type of float give from 6 to 9 significant decimal digits while double gives 15–17 digit precision, because of the way binary translates to decimal. In this case seven decimal digit are adequate, so a function or a preprocessor macro such as #define EZERO(v) ((v)<(0.0000005)) is called for.

Looking closely at Fig. 8 it is apparent that we may not necessarily need to check every point or line. Triangle D has two points inside triangle A, so the first one we find is sufficient to know they overlap. Similarly, all three sides of F intersect sides of A, so a single intersection test would be sufficient. B and E have no points inside A, so only the intersection test will offer insight, but C has no intersections with A so the point test would quickly confirm if it is completely inside or completely outside. There is no way to tell in advance if the point or intersection test will be more efficient, although with the intersection test it might be possible to increase the odds of not choosing the wrong line by checking closer line segments first. Depending on the power, processor and memory resources available, it could be possible to put the triangles or the line segments into a spatial tree structure of some type - a grid, quad-tree or kd-tree would allow testing multiple triangles or multiple line segments simultaneously.

## 5. Results and Discussion

As noted above, the goal is for a node to establish an initial estimate of its location and a finite set of alternate points that could be its actual location, given minimal information. Given the coordinates and transmission radius of two or three neighbors, this can be done by passively listening to transmissions and doing some efficient calculations to determine the farthest a node can be from its neighbors and still receive transmissions. As more nodes join the network and go through this process more information becomes available, and the nodes can narrow their set of possible actual locations through a process of iterative refinement. Ultimately the nodes in the network will be able to converge on a stable network map within a quantifiable margin of error for each node.



The overall accuracy of this network map will depend upon several factors. First, the mathematical process of iterative refinement will require decision rules for choosing amongst several possible locations, and each choice will have a "ripple effect" as the data is used by other nodes. For there to be any relationship to reality at all, some nodes must know their coordinates with complete confidence (ground-truth) to seed the network map, e.g., having a GPS receiver or preprogrammed with accurate GPS coordinates. A node with more than one ground-truth neighbor should be able to interpolate its location with a high degree of confidence (possibly even accuracy), while nodes with no ground-truth neighbors will necessarily have lower confidence in their position estimates. However, as Table 1 shows, there would be an upper limit on accuracy of 8-10 meters using a typical GPS unit with no interference.

The additional information of whether an audible or inaudible neighbor is a ground-truth node or has a ground-truth node as a neighbor would be useful for decisions about where to move within the set of alternative locations. These nodes are the best starting point for the process of determining the overlap of the triangles that define sets of alternative locations. Staying within the union of these sets will yield an estimate that can be considered more accurate than any potential location calculated with data from neighbors that do not have direct or indirect knowledge of ground-truth. A simple move to the middle strategy might be enough to provide acceptable accuracy when these circumstances apply.

The transmission radius of a node is essential information required for determining relative location, but this is also a source of uncertainty. This information can be configured, but will always be a best guess because of the numerous factors that affect path loss. The terrain over which signals travel, the level of moisture in the air, the shape of obstacles and their location relative to the two antennas can all affect signal reception, individually or in concert. In practice, models based on empirical measurements over a given distance in a given frequency range for a particular geographical area or building are used to describe signal propagation, recognizing that these represent an average and are less accurate in a more general environment. Further, empirical measurements have shown that the difference between the average and the actual path loss is random and log-normal distributed, which means that any receiver in range of the transmitter has a nonzero probability of receiving a signal that is too weak to use, and some nodes beyond the average range will receive a usable signal [6].

The practical implication is that the value chosen for the transmission radius when configuring the network nodes must be considered nominal. If the case depicted in Fig. 8 arises, where no single point satisfies all of the constraints, the overall accuracy of the network map should be improved by selecting a point that takes "reasonable" variation into account. It might also be useful to extend the capabilities of the nodes to communicate an indicator of the average quality of the signal received from each



audible neighbor, so the decision rules could be tuned to prefer to move closer to stronger signals and not move farther away from weaker sources.

## 6. Conclusion

It is possible to create a system where a WSN node can calculate an initial estimate of its location and a finite set of alternate points that could be its actual location, given the coordinates and transmission radius of two or three neighbors. The necessary information can be acquired by passively listening to transmissions, assuming a nodes in the network can transmit their actual or presumed location and nominal transmission radius; optimally the nodes would also be able to communicate the location and transmission radius of its audible and inaudible neighbors, and the average quality of the signal received from each audible neighbor. Exactly how this is achieved (e.g., through beaconing, addressing, or some type of protocol) is not important for the calculations.

As with all WSN localization techniques, a primary goal is to derive a satisfactory degree of accuracy from inconsistent radio communication while minimizing power consumption. The set of functions presented here provide efficient calculations to determine the farthest a node can be from its neighbors and still receive transmissions. Three base cases (two audible neighbors, one audible and two other neighbors, one audible neighbor that has only one audible neighbor) are sufficient for the initial calculations. The process of narrowing the set of possible actual locations through iterative refinement as more nodes join the network is where the decision rules will have to be tuned for specific characteristics of the deployment, in order to determine a quantifiable margin of error for each node. The result is a unique low-cost way to address limitations on determining directionality in broadcast networks.

#### Acknowledgments

Development of the functions presented here has benefitted immensely from innumerable discussions and suggestions posted in various StackExchange fora (<u>http://stackexchange.com/</u>).

#### References

- [1] V. Gungor, and G. Hancke, "Industrial Wireless Sensor Networks: Challenges, Design Principles and Technical Approaches", *IEEE Trans. Ind. Electron.*, vol. 56, no.10, 2009.
- [2] L. Cheng et al., "A Survey of Localization in Wireless Sensor Networks", *Int. J. Distributed Sensor Networks*, vol. 2012, Article ID 962523, Dec. 2012.



- [3] M. Srbinovska, et al, "Localization Techniques in Wireless Sensor Networks using Measurement of Received Signal Strength Indicator". Electronics, vol. 15 no. 1, 2011.
- [4] H. Maheshwari, "Optimizing Range Aware Localization in Wireless Sensor Networks", PhD Dissertation, Fac. Eng., Univ. of Leeds, UK, 2011.
- [5] Sandor Szilvasi, et al, "Interferometry in Wireless Sensor Networks", in Interferometry - Research and Applications in Science and Technology, Rijeka, Croatia: InTech, 2012, ch.21, pp. 437-62.
- [6] M. Tsai, (2011, Oct.20) "Path-loss and Shadowing (Large-scale Fading)." National Taiwan University [Online]. Available: <u>https://www.csie.ntu.edu.tw/~hsinmu/courses/\_media/wn\_11fall/path\_loss\_and\_shad\_owing.pdf</u>.
- [7] CK-12 Foundation, (2017), 2.1 Radian Measure [Online]. Available: http://www.ck12.org/book/CK-12-Trigonometry-Concepts/section/2.1/?noindex=None
- [8] Florida Center for Instructional Technology, (2017), *Radians in Complete Circle*, [Online]. Available: <u>http://etc.usf.edu/clipart/36600/36685/radians\_36685.htm</u>
- [9] P. Osborn, *The Mercator Projections*, Zenodo, 2013. doi:10.5281/zenodo.35392
- [10] D. Williams, (2016, Dec. 23) *Earth Fact Sheet* [Online]. Available: <u>https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html</u>
- [11] T. O'Daniel et al., "Localization using GPS Coordinates in IPv6 Addresses of Wireless Sensor Network Nodes", *Indian J. Sci. & Technology*, vol. 10, no. 8, Feb. 2017.
- [12] W. Huber, (2011, Apr. 18) "How to measure the accuracy of latitude and longitude?" [Online]. Available: <u>http://gis.stackexchange.com/questions/8650/how-to-measure-the-accuracy-of-latitude-and-longitude/8674#8674</u>
- [13] E. Williams, (2011, Apr.24) *Aviation Formulary v1.46* [Online]. Available: <u>http://edwilliams.org/avform.htm</u>
- [14] C. Veness, (2015) *Latitude/longitude spherical geodesy formulae & scripts* [Online]. Available: <u>http://www.movable-type.co.uk/scripts/latlong.html</u>
- [15] P. Bourke, (1997, Apr) "Intersection of Two Circles", *Circles and Spheres* [Online]. Available: <u>http://paulbourke.net/geometry/circlesphere/</u>
- [16] R.Chamberlin, (1996, Oct.) "Q5.1: What is the best way to calculate the great circle distance (which deliberately ignores elevation differences) between 2 points?", *Geographic Information Systems FAQ* [Online]. Available: <u>http://www.faqs.org/faqs/geography/infosystems-faq/</u>



# A Review on Social Media Issues and Security Awareness among the users

Nor Azlina Abd Rahman Faculty of Computing, Engineering & Technology Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia nor\_azlina@apu.edu.my

Fauziah Permatasari Faculty of Computing, Engineering & Technology Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia fauziah.p@mail.com

Yuni Hafsari Faculty of Computing, Engineering & Technology Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia yunihafsari158@gmail.com

*Abstract* - This paper is reviewing on several issues faced by the users of social media especially users of Facebook and Twitter which are the most popular social media networks based on statistic provided by Statista. Some of the issues that being discussed are information theft and image misused due to no privacy setting set by the users to their social media accounts. The use of social media aggregator will lead to the hackers to easily access to all the Social Medias that are link together. Lack of security awareness among the social media users will exposed them to many cyber-crime activities. Several countermeasures are being discussed such as to improve the privacy setting of the social media accounts are by disable the connection between the infected accounts with other accounts by changing the registered email information on the unaffected accounts. Changing passwords could help to slowdown the hacker's activities and the last option that can be considered is by reporting and issues the social network's customer service

Index Terms – cyber security; social media; facebook; twitter; wikipedia; phishing



## 1. Introduction

**S**ocial media and its user has a relationship that is affected with cyber security and its environment. As the social media use explodes, this relation also becomes more intense as cyber criminals expand their hunting area and start to have their eyes on the social media accounts. Social media users are becoming one of their targets as most of the social media users are lack of knowledge and awareness in security and privacy that can be implemented to individuals account.

Social media nowadays have played a crucial part of people's daily life. Based on Fig. 1 showed that in 2016 Facebook had over 1,590 million active users while the Facebook-owned WhatsApp and Instagram had 1,000 million and 400 million respectively [1]. This clearly proved that social media has become more popular nowadays and this further proves that Social Media users have reached a significant number to be influential, and makes them even more vulnerable as hacking victims.

In this paper, Facebook and Twitter users are the main focus of discussion, considering that these two social networks are wide-spreading rapidly, as proven by the GlobalWebIndex Wave 11 that reported Facebook and Twitter have the highest penetration rates with 93 percent of Internet users owning a Facebook account, and 72 percent of them are a Twitter account holder [2].



Fig. 1: Survey result of leading social media users as of April 2016 [1].

In relation with the rise of Internet popularity around the world, it especially has become even more popular amongst teenagers [3]. Based on Fig. 2, Social media users'



number have significantly risen for the past few years both for adult and teen users. In 2008, 73% of American teens had been using social networking websites, an increase of 20% compared to the survey on 2006 [4].

Most of these teenager social media users are students whom the global environment changes had significantly affected them both academically and socially. These teenagers and young adults are clueless regarding their privacy settings. Since teenagers and young adults are more concerned on relishing the opportunity to link themselves to others and create authentic relationships, they want to express their identity and take the risk of exposing themselves to being discovered and come into contact with hackers [5] [19].



Fig. 2: Graph of American teenagers, young adults, and adults' use of social media [4].

## 2. Issues faced by social media users

In terms of personal security, users have done several activities that put them face to face with the risk of being a cyber-crime victim. Some of these social media users are digging their own grave in the case that they are exposing themselves toward the cyber-crimes and privacy leakage risks. The activities that they did on social media may seem trivial when actually it is significant in helping them to be one of a hacker's victims [6]. There are several issues that faced by the social media users which are:

#### 2.1 Lack of social media privacy setting

Most of the users of social media are not really aware the importance of their social media privacy setting. Teenagers and young adults, are the most prone of being nonchalant regarding their social media privacy setting. According to Livingstone's research result (2008), teenagers' main use of social media is taking risky opportunities in youthful





content creation by expressing themselves and intimacy. One example of a cyber-crime that is resulted by this lack of privacy setting is identity thievery. A related case as reported by The Telegraph [7], a 27-year-old woman in London, UK, had her Facebook profile pictures and identities stolen and they were being used as a fraud in an online dating social media. This gave her many social disadvantages as it would stain her professional reputation along with her relationship.

#### 2.2 The use of social media aggregator

In relation to the first point, as the number of social media users grows, then so does the percentage of social media users who actively use different social network sites. This leads to the second activity on which users are easing hackers to discover them. Since the users are maintaining multiple social networks for their personal uses, they often link each of the social network account authorization to each other, or use the same password for different account [8].



Fig. 3: Example of how a user links their social media accounts altogether [9].

As for the hackers, this can be said as heaven-sent target. They can easily gain access to multiple account just by gaining access to one of the person's many accounts. This is called as social network aggregator. Fig. 3 showed how the social network aggregator works when a platform is used to pull content from multiple social media into a single location and dispersing it to different profiles more conveniently. Although it is eases users on monitoring their social media accounts, it is relatively unsafe [9]. Considering hackers will be able to discover all of their other accounts once one of them has been compromised, putting other accounts on huge risk.

One example of this is the latest case with Facebook's founder, Mark Zuckerberg. A group of hackers has successfully gained access to his Twitter and Pinterest account using his LinkedIn credentials which had been leaked in a prior huge data loss of LinkedIn user information [10]. Even a person who is as sophisticated as Mark Zuckerberg in information technology and cyber industry, is a victim of such hacking activity. It opened many people's eyes on how vulnerable their social media accounts are.

#### 2.3 Insufficient information about cyber crime

Insufficient information about cyber security can also be an issue to these teenagers and young adults. Due to the unawareness of the importance of the security implementation,



makes them unaware of incidents of accounts being compromised which are happening around them.

Social engineering attacks may not sound as sophisticated as other methods of hacking, but in fact, it has some of the most successful attempts carried on targets. A report by Symantec (2016) showed that some of the attacks and tactics executed by these cybercriminals have deliberately proven how vulnerable the Internet users are online.

Social media especially took the number one spot of scamming target, as criminals seek to gain people's trust by having them in their social circles therefore these hackers would be able to spread scams, fake links, and phishing. An example of this case is Mocking birds, which is where thousands of fake Twitter accounts are made to gain branch boosting, get followers, and having retweets from legitimate Twitter users. This creates even more possibility of phishing cases to happen. While according to BofA Merrill Lynch Global Research, cybercrime already costs the global economy up to US\$575 billion annually [11].

On December 21, 2016, U.S. Netflix's Twitter account was hacked by OurMine, a hacker group which has been known for hacking into several high-profile social media accounts and popular websites [12]. Some of its victims include Twitter and Vine co-founder Google chief executive Sundar Pichai, Wikipedia co-founder Jimmy Wales, Forbes, Buzzfeed, and Techcrunch. These people are in no way amateurs in information technology and its cyber security. However, in the latest case on which Netflix was compromised, OurMine shared that it gained the access of some Netflix accounts by hacking into its Director of Social Media who apparently had not changed its LinkedIn password post the recent user information leakage [13] [20].

These little things that some people may find as insignificant, can lead them to a chain reaction and result in a big loss. Netflix gains people's frowns and questions on its take on security and safety of its users' privacy. Because social media is a persona that people and organizations want the industry and the society to view them as. Social media has become a part of people's daily life; it now holds more worth than how it used to be just ten years ago, a fact that definitely is noticed by irresponsible parties such as hackers.

## 3. Countermeasures

In this section, countermeasures are discussed to avoid and prevent exploitations on personal social media accounts. Besides that, the recovery plan for attacked accounts will also be suggested in order to prevent further damage. Several countermeasures that can be considered are as the following:

#### 3.1 Improve privacy setting

There are numerous security implementations that social media users can do to prevent cases mentioned above. For the first point on which social media users are not paying



attention toward their privacy settings, countermeasures can be done both on their social network sites and their internet browsers.

Nowadays social networking sites have sufficient privacy settings which its users can adjust and customize based on their level of comfort. Users can set some of their identity information on private to be viewed only by selected users. As an example, is Facebook that upgraded its privacy profile privacy setting [14] to enhance the security.

Facebook allowed people to manage their privacy settings in different granularities, making it easier for the users to choose to whom they are sharing their profile content with. Even so, users are advised to keep their awareness on everyone because hackers can also be someone they are close with. Therefore, another tip that need to consider to ensure the security of the social media users are by avoiding sharing everything of their daily life to the internet [15].

Meanwhile for their browser, users can set its privacy setting to prevent the use of cookies and block pop-ups. In several browsers, they also have a location setting which the user can turn off too. This will lessen the risk of users to be exposed to much more contact with irresponsible parties.

#### 3.2 Control social media authorisations

The users of social media should not use any social network aggregator or not to connect their social media accounts altogether. It may seem difficult to some social media users who already had their social network accounts linked to each other. Nevertheless, they can still set a different password for each of their social media account to decrease the risk.

The passwords used for each social media should be strong, meaning that it should consist more than six characters with upper case letters and numbers mixed in it. Make sure that the password is not of a common word or saying. This can prevent hackers to crack a users' account. Hackers are able to crunch and automatically run a tool which can generate a dictionary of commonly used phrases for passwords to crack a social media account.

#### 3.3 Increase awareness of latest cyber security issues

Social media users have to kick their bad habits and start to safeguard against social engineering and even the cybercrime in general. They have to be more aware of the cybersecurity events and news. Also, they have to implement the good cybersecurity basics starting by themselves.

For the Twitter users, they have to be more cautious in choosing the people they want to follow or they want to approve of. They have to be sceptical for each of their new followers, to prevent of getting spammed by bots. And they also have to pay more attention on the 'verified' badge when they want to follow certain accounts.


# 4. Recovery Plan

It is very important to have a recovery plan as it will cover the steps to take after the social media has been hacked to prevent any more loss from happening. Specifically, for an individual whose professional work revolves around their social media use, which as stated before, their social media accounts are a persona to them, they need to consider on cooperating with an expert of the field. How to do the damage control and recover from such a state [16].

The crisis shall be divided into several levels depending on the damage impact. Every impact will require different response action, based on the situation. Some crisis communication exercises and scenarios can also be planned within the recovery plan, in order to help the crisis handlers to be ready whenever a crisis happens [17].

One example of things to consider listed in personal social media protection is by taking an immediate action to change other social media accounts' passwords. This relates to the social network aggregation in which when a hacker has gained access to one of a user's accounts, he might have access to other accounts as well. Changing other accounts' passwords will help minimizing the risk.

Then, if it is possible, disable the connection between the infected social accounts with other ones. This can be done by changing the registered email information on the unaffected accounts, and users are also encouraged to report the case to the social network's customer service.

# 5. Conclusions

Modern technologies help us a lot in communication nowadays. Based on research and statistic that has been discussed above, social media become one of the popular communication media among teenagers and young adults. Lack of security awareness among the users lead to many security issues such as information theft, image misused and others. Hence the users have to pay more attention on how they access and manage their accounts [18]. This is quite often when the little things which are insignificant to them, may lead into a much bigger reaction which can put them in risk. They also have to be in touch with the latest news and findings about information and cyber security issues. By being update regarding those, the social media users can adapt themselves to the situation and be more knowledgeable on cyber security and protect themselves. Social media is part of people's image in the society nowadays and a compromise in these accounts can cause damages toward the victims.

# Acknowledgments

The authors wish to express gratitude to the management of Asia Pacific University for their support.



# References

- [1] Statista, (2017). The most popular social networks worldwide [#ChartoftheDay] -Smart Insights Digital Marketing Advice. [online] Smart Insights. Available at: http://www.smartinsights.com/digital-marketing-strategy/popular-socialnetworks-worldwide-chartoftheday/ [Accessed 8 Jan. 2017].
- [2] Global Digital Statistics, (2014). GWI Social Summary. GWI Quarter Report. [online] Global Web Index. Available at: http://insight.globalwebindex.net/hsfs/hub/304927/file-2377691590pdf/Reports/GWI\_Social\_Summary\_Q4\_2014.pdf?submissionGuid=d75c46ce-922c-4efc-8ac3-08dbe9ba4904 [Accessed 1 Feb. 2017].
- [3] Bennett, S., Bishop, A., Dalgarno, B., Waycott, J. and Kennedy, G. (2012). Implementing Web 2.0 technology in higher education: A collective case study. 1st ed.
- [4] Lenhart, A., Purcell, K., Smith, A. and Zickuhr, K. (2010). Social Media & Mobile Internet Use among Teens and Young Adults. In: ERIC, 1st ed. Pew Internet & American Life Project.
- [5] Livingstone, S. (2008). Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. New Media & Society, 10(3), pp.393-411.
- [6] Marwick, A. and Boyd, d. (2014). Networked privacy: How teenagers negotiate context in social media. New Media & Society, 16(7), pp.1051-1067.
- [7] Gammell, K. (2017). My Facebook profile was stolen to get dates on Tinder and there's nothing I can do. [online] Telegraph.co.uk. Available at: http://www.telegraph.co.uk/women/womens-life/11588667/Facebook-identity-theft-My-profile-was-stolen-to-get-dates-on-Tinder.html [Accessed 8 Jan. 2017].
- [8] Lewis, K., Kaufman, J. and Christakis, N. (2008). The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. Journal of Computer-Mediated Communication, 14(1), pp.79-100.
- [9] Benevenuto, F., Rodrigues, T., Cha, M. and Almeida, V. (2012). Characterizing user navigation and interactions in online social networks. Information Sciences, 195, pp.1-24.
- [10] Hern, A. (2017). Mark Zuckerberg hacked on Twitter and Pinterest. [online] the Guardian. Available at: https://www.theguardian.com/technology/2016/jun/06/mark-zuckerberghacked-on-twitter-and-pinterest [Accessed 8 Jan. 2017].
- [11] Symantec, (2016). 2016 Internet Security Threat Report. Internet Security Threat Report. [online] Symantec Corporation. Available at: https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016en.pdf [Accessed 1 Feb. 2017].



- [12] BBC News. (2017). Netflix US Twitter account hacked BBC News. [online] Available at: http://www.bbc.com/news/technology-38390343 [Accessed 8 Jan. 2017].
- [13] Schroeder, S. (2017). Hacker group OurMine has compromised another high-profile social media account. [online] Mashable. Available at: http://mashable.com/2016/12/21/netflix-twitter-hacked/#7BnKhHU70mqE [Accessed 8 Jan. 2017].
- [14] Liu, Y., Gummadi, K., Krishnamurthy, B. and Mislove, A. (2011). Analyzing facebook privacy settings. Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference - IMC '11.
- [15] Harkins, M. (2013). Managing risk and information security. 1st ed. [New York]: Apress.
- [16] Pensa, R. and Di Blasi, G. (2017). A centrality-based measure of user privacy in online social networks. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), (2016-08), pp.1438-1439.
- [17] Ryan, D. and Jones, C. (2012). Understanding digital marketing. 1st ed. Philadelphia, PA: Kogan Page.
- [18] Chu, S. (2011). Viral Advertising in Social Media. Journal of Interactive Advertising, 12(1), pp.30-43.
- [19] Charlesworth, A. (2015). An introduction to social media marketing. 1st ed. London [England]: Routledge, p.209.
- [20] Cross, M. (2013). Social Media Security: Leveraging Social Networking While Mitigating Risk. 1st ed. Newnes.



# An Insight into Programming Paradigms and Their Programming Languages

M. Selvakumar Samuel Faculty of Computing, Engineering & Technology Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia dr.selvakumar@apu.edu.my

**Abstract** - A Programming Paradigm is the silent intelligence in any software design. Although many Programming Paradigms have evolved, only a few programming paradigms are actively used by the software industry. In addition, many hundreds of programming languages have been developed, but only a few are established and beneficial. The main aim of this paper is to provide an in-depth view into this area in order to give an opportunity for the Academia, Researchers, and the Software Industry to understand this domain in a different way. Basically, in this paper, a lot of relevant literatures have been reviewed and some useful facts, such as mainstream programming paradigms, suitable programming languages for the current software development scenario, weaknesses in the current research works in this domain, etc., have been derived as conclusions. The deduced facts would be beneficial for the education sector to decide the programming paradigms and programming languages to teach at this juncture, and as for the researchers, this paper would provide an alternative road map to conduct further research in this domain. Eventually, this work would benefit the software designers to choose appropriate programming paradigm concepts and their respective programming languages based on the deduced facts as the result of this study.

*Index Terms* - Programming paradigm; programming language design; software development

# 1. Introduction

A programming paradigm is the core and basis for any software and programming language design. There are many accepted definitions for the term "Programming Paradigm". According to Daniel [1], a programming paradigm is "a style of programming expressing the programmer's intent". Linda [2] said that, "it is an approach to solving programming problems" and Pamela [3] said that, "it is a way of thinking about computer systems".

In software development, there are many paradigms [4] that have been suggested and these paradigms keep evolving in order to suit the requirements of the software development of the respective times. However, only a few programming paradigms, such



as Imperative, Object-Oriented, Functional, Event Driven with Graphical User Interface (GUI), Logic, and Concurrent, are widely accepted or being used by the software industry. Amongst these few programming paradigms, the Object-Oriented programming paradigm can be considered as the dominant programming paradigm.

Each programming paradigm exclusively has its own approach, purpose, merits and demerits. Similarly, each programming paradigm has its own set of programming concepts. When designing software with the chosen programming paradigm environment, the respective programming paradigm concepts will be used. For example, when designing our solutions purely in the imperative way, Imperative programming paradigm concepts, such as control structures, input/output statements, assignment statements, etc., will be used.

A programming language is actually a collection of libraries or API's, which are based on a core programming paradigm and supports some other programming paradigms as well. For example, C++ is basically an Imperative programming language, but it supports Object-Oriented and others.

The main objective of this paper is to critically evaluate the programming paradigms adopted in programming languages with respect to the current software development context. The outcome of this paper is basically to provide an idea to the academicians to decide on the appropriate programming paradigms and programming languages to be covered based on the recent trends and current requirements. This work also attempts to provide a roadmap to the researchers working in this area and also to provide support for the software industry, especially in the stage of the software design phase of the software development life cycle.

The following contents evaluate the Programming Paradigms in terms of various works in literature whereby enabling the deduction of the relevant facts in order to achieve the above said objectives.

# 2. Most Influenced (Mainstream) Programming Paradigms

Every single programming language is based on one or more programming paradigms. Each programming paradigm consists of a set of programming concepts [5]. There are a huge number of programming languages, but only 27 different programming paradigms are being used [6]. Amongst these 27 programming paradigms, only a few are actively being used by the software designers.

The IEEE Spectrum [7] has ranked the programming languages based on 12 metrics across 10 reliable sources. Over 48 programming languages were analysed with a number of different dimensions. The results are summarised in Fig. 1 and Fig. 2.

Table 1 shows the list of the top programming languages and their respective programming paradigms. In this list, most of the languages are mainly realised from the programming paradigms, such as Object-Oriented, Imperative, Event-Driven with GUI, and Functional.



Lan	guage Rank	Types	Spectrum Ranking (2	2015) Spectrum Ranking (2014)
1.	Java		100.0	100.0
2.	с		99.9	99.3
з.	C++		99.4	95.5
4.	Python	• -	96.5	93.5
5.	C#		91.3	92.4
6.	R	Ţ	84.8	84.8
7.	PHP		84.5	84.5
8.	JavaScript	⊕ □	83.0	78.9
9.	Ruby		76.2	74.3
10.	Matlab	Ţ	72.4	72.8

Fig. 1: Top 10 Programming Languages in 2014 and 2015 [7]

Language Rank Ty		Types	Spectrum Ranking		
1.	С		100.0		
2.	Java		98.1		
з.	Python		98.0		
4.	C++		95.9		
5.	R	$\Box$	87.9		
6.	C#		86,7		
7.	PHP	•	82.8		
8.	JavaScript		82.2		
9.	Ruby		74.5		
10.	Go		71.9		

Fig. 2: Top 10 Programming Languages in 2016 [8]

#### Table 1: Top programming languages in 2016 and their respective programming

putuugino				
Number	Programming	Main Programming Paradigm(s)		
	Language			
1	Java	Object-Oriented, Imperative, Event-driven with GUI, Concurrent,		
		Functional, Generic, & Reflection		
2	С	Imperative (Procedural and Structural)		
3	C++	Imperative, Object-Oriented		
4	Python	Imperative, Object-Oriented, Functional, Event-Driven with GUI,		
		Concurrent, Reflection, & Meta programming		
5	C#	Object-Oriented, Imperative, Event-Driven with GUI, Functional,		
		Concurrent, Generic, & Reflection		
6	R	Functional, Object-Oriented, Event-Driven with GUI, Imperative,		
		Reflective, & Array		
7	PHP	Imperative, Object-Oriented, Event-Driven with GUI, Functional, &		
		Reflection		
8	Iava Script	Scripting		

paradigms



9	Ruby	Functional, Object-Oriented, Imperative, Event-Driven with GUI, & Reflection	
10	Go	Concurrent, Logic, Functional, & Object-Oriented.	

Zuhud et al. [9] state the main programming paradigms as Functional, Imperative, Object-Oriented (OO), and Logic. Souza et al. [10] assert that, in today's current software development, the most widely used programming paradigms are Object-Oriented and Procedural. Similar to these researchers, many other researchers, academicians, books [11,12,13,14,15,16,17,18,19,20,21, &22], & literature works mainly discuss Imperative, Object-oriented, Functional, and Event-Driven with GUI programming. Event Driven with GUI is common in every current software. Particularly, every current software's interactive input and output designs are based on Event Driven with GUI approaches. Apart from this, many industrial, open source and academic software projects have been investigated. It is found that, these four mainstream programming paradigms are dominant in all software projects.

These findings fairly conclude that, these four programming paradigms are the most popular or mainstream programming paradigms in the industry, as well as in the academic domain. Amongst these, the Object-Oriented programming paradigm can be considered as the dominating paradigm; whilst the Functional programming paradigm is the emerging programming paradigm. Hence, in this research, Imperative, Object-Oriented, Functional, and Event Driven with GUI have been considered, and the research works which are related to these four programming paradigms and their concepts have been examined in the following sections.

As these four programming styles mainly dominate the software industry and the academic domain, the ideologies and the programming concepts behind these programming paradigms are obvious and very familiar. Hence, only a brief introduction of these four mainstream programming paradigms are stated as follows:

- Philip Roberts [23] states that in Imperative programming, the user instructs the computer as to what the user wants, and also instructs the computer as to how to get what the user wants. In other words, the user needs to define the details, step by step, for the computer in order to reach the goal. This type of programming is also known as algorithmic programming [24]. Procedural and Structural are the common Imperative style programming paradigms [25]. The Procedural programming paradigm is a traditional programming approach and it is the basis for the CPU's fetch-decode-execute cycle, as well. To produce the desired results, this programming paradigm has programs defined as a sequence of instructions which manipulates data to output the desired results [26].
- Object-Oriented programming is an engineering approach for building software systems which are based on concepts of classes and objects that are used for modelling the real-world entities, which changes the focus of attention from code to data. The general idea of object technology must be represented in the Object-Oriented programming languages so that complex problems can be solved in the same way as real-world situations.



- The idea of Functional programming was initially influenced by Alonzo Church's lambda calculus, which uses mathematical functions as the basic building blocks of the system [27]. Just as the name implies, a functional program is made up of functions which comprise the definition of the expressions and these expressions will be evaluated during program execution to yield the final results [28]. Hence, the main program itself is just like a single function which contains several operational computations. Programmers do not need to worry about its computational complexity as all of the complex operations will be laid in the back-end of the machine [29]. This makes functional languages more expressive.
- Event-Driven programming and graphical user interfaces (GUIs) are all interrelated [30]. The graphical interface objects or components on the forms, windows, or containers make up the view and control of an application, according to the Model-view-controller (MVC) framework [31]. Unarguably, GUI systems are Event-Driven oriented and connote that the system program does not flow sequentially from the start to the finish. "Event-Driven" literally means to be interrupted by an event or driven by an event; hence, waiting for something or an event to happen before the appointed response to that event occurs. GUI programming is among the trickiest programming paradigm shift [30].

Some other programming paradigms, such as Reflection, Concurrent, Generic, Array, Data Flow, Meta, and Constraint are also supported by these programming languages, but they do not have as rich a set as the programming concepts as the mainstream programming paradigms. Generally, these non-mainstream programming paradigms are either a technique to improve the program behaviour or a programming ability or a method to solve a specific problem, and they are always working very compatible with the mainstream programming paradigms.

In order to understand the nature of these non-mainstream programming paradigms, a brief introduction is stated as follows:

- DataFlow programming (DFP) internally represents applications as a directed graph, similar to a Dataflow diagram [32].
- Array programming mainly helps programmers in designing complex data analysis procedures [33].
- Reflection is the ability of a program to manipulate data as something representing the state of the program during its own execution.
- The kind of change that can be performed in the meta-programming approach is to modify the meta-interpreter prior to the execution of the program. Reflection, on the other hand, allows the program to change its behaviour whilst running, depending upon its current execution (such as the inputs and intermediate results) [34].
- Generic Programming is based on the principle that software can be decomposed into components which make only minimal assumptions about other components, allowing the maximum flexibility in the composition [35].



- The basic idea in constraint programming is that the user states the constraints, and a general-purpose constraint solver is used to solve them [36].
- A Concurrent program is the one defining actions that may be performed, simultaneously [37].

Amongst these non-mainstream programming paradigms, concurrent programming has relatively vast application areas.

# 3. Programming Paradigm Notions / Concepts

A programming paradigm is a generic way to design a program. Each programming paradigm has its own set of programming concepts. Class, Object, Inheritance, Data Hiding, etc. are called the programming notions or concepts of the Object-Oriented Programming Paradigm. In order to design a programming solution in the Object-Oriented way, the object-oriented programming concepts will be used.

Most of the concepts have their equivalents, and they produce the same results. For instance, Imperative programming iterative control structures are: for...loop, while...loop, and do...while...loop. These three control structures are alternatives to each other and they produce, generally, the same results.

Event-Driven with GUI programming is different from other programming paradigm concepts. As for the Event-Driven with GUI programming, GUI controls/elements are the key notions of event handling mechanisms in each development platform. Different elements of a graphical user interface exist on different platforms, including on modern day smartphone platforms, such as Android, IOS, and Windows. However, there is also a large overlap amongst the GUI elements that are common to all mobile operating systems (and even desktop operating systems) [38]. On a given User Interface (UI), some of these elements, generally or under certain circumstances, can be replaced with other similar elements. For example, the day of the week can be selected using a dropdown menu or radio buttons.

Application developers often apply these concepts without paying much attention to the impact they might have on the output of an application [39]. The concepts are the basic primitive elements used to construct the paradigms [40] and programming languages. In this section, some of the studies which predominantly evaluated the efficiency of the Programming Paradigms and their concepts are explored. Other related studies are discussed in the later sections.

#### 3.1 Imperative Programming and its Concepts

The core concepts of the Imperative programming paradigm are Control Structures (looping constructs and iterations), Input /Output concepts, Error and exception handling, Procedural Abstraction, Expressions and assignment statements, and Library support for data structures [41]. Some relevant works have been identified in this area.

Oracle has studied String Builder with String Buffer as both are performing the same operations, but utilizing different volumes of resources. The result of this study



proved that String Builder consumed less resources than String Buffer [42, 43]. Likewise, they have studied ArrayList with VectorArray. They have found that the ArrayList performed better than VectorArray [44,45].

Most Imperative Programming languages support both recursion and iterations. Rubio-Sánchez [46] has said that, recursion is a key concept in computer science and mathematics, and it is a powerful problem-solving tool, which constitutes an attractive alternative to iteration, especially when problems can be solved using a divide and conquer approach. Imperative programming uses Ad-hoc recursion whilst functional programming has introduced the tail recursive approach. In languages that favour iterative looping constructs, there is usually significant time and space cost associated with the recursive programs due to the overhead required to manage the stack and the relative slowness of the function calls. Recursion and iteration are used to solve a task one at a time and finally, combine the results. Iteration emphasises repeating a task until it reaches its counter limit. Recursion emphasises breaking a larger problem into smaller pieces until the problem is solved. Looking at the efficiency point of view, recursion calls the same function over and over; whereas, iteration jumps to the beginning of the loop. The function call is normally more expensive than the jump.

Er [47] has evaluated the performance of recursive and iterative algorithms in terms of their time and space requirements. The evaluation was performed on the programs written for the famous game Towers of Hanoi. The result showed that the iterative approach outperformed the recursive one in both time and space performance. In a popular study, Schaeckeler & Shang [48] suggested that if a formal parameter or local variable is dead at all recursive calls, then it can be declared globally, so that only one instance exists independent of the call depth. The research also found that in 70% of popular recursive algorithms and in all our real-world benchmarks, it is possible to reduce the stack size by declaring formal parameters and local variables, globally. The stack size reduction starts to materialise for their benchmarks no later than in the fifth recursion on a 32-bit Intel Architecture.

Venkat Subramaniam [49] has conducted an experiment to perform a Factorial calculation on Java VM using Iteration, Recursion, and Tail Recursion with the Scala language. When computing the Factorial value of 10000 using Recursion, he got a Stack Overflow Error, but the Factorial value of 5 computed just fine. Furthermore, the iterative version of Factorial value of 10000 was able to be computed without any problem. He pointed out that the iterative version made use of a mutable local variable that the recursive solution nicely avoided. The best of both worlds could be achieved, if a recursive code using a compilation technique could transform the code and run as an iterative process.

In another study, Liu & Stoller [50] explain that transforming recursion into iteration eliminates the use of stack frames during the program execution. In a study by Schaeckeler & Shang [51], it was shown that compilers tried to reduce the code segment and neglected the stack segment; although, the stack can significantly grow during the execution of recursive functions.



Likewise, the study conducted by Chandran [52] to test the implementation of the Cyclomatic Complexity analysis feature in Coverlipse(an open-source coverage analysis tool), the researcher examined the program written in a *switch* statement and three other programs written in different constructs of *if* statements, namely, if-else statements, if statements, and an un-indented if-else statement. His research showed that these alternative styles of implementing a control flow program had resulted in different numbers of cyclomatic complexities, which meant that they had different levels of software complexity. This result further proves that both conceptual and syntactic level evaluations are feasible for performance evaluations.

Most of the researchers in the Imperative programming area are mainly focussing on the efficiency of the control structures and the recursive function call features.

# 3.2 Object-Oriented Programming (OOP) and its Concepts

The benefits of Object-Oriented programming include reduced complexity, object re-use, enhanced modularity, encapsulation, design benefits, and software maintenance. However, the OOP paradigm can increase power consumption and execution time, thereby hindering performance. Mattos, C.B. and Carro, L., [53] highlighted that Object-Oriented programming significantly increases dynamic memory use, thereby developing overhead in terms of memory, performance, and power. Due to various reasons, there are only a handful of documented bad experiences with object technology.

Boasson, M. [54] stated that companies are not very forthcoming with details of failed projects and secondly, it is difficult to get anything published that questions the usefulness of OOP. Researchers, such as [55, 56, 57, & 58], have demonstrated the consequences on the object-oriented system performance, in terms of execution time and memory overhead.

Kayun, C. and Chonawat, S. [59] have carried out a research which aims to find appropriate methods and recommendations of using some OOP concepts that consume less power by using real quantitative results. This research has focussed only on power consumption, but this is the only work targeted to identify the optimised concepts of Object-Oriented programming paradigms.

Several researchers have reported the issues and other concerns on the Object-Oriented programming paradigms, but very few works have been found related to the Object-Oriented programming notions. Some of the OOP concepts have alternative choices which can be used instead of the others, in some cases. Due to the differences in these concepts, some are lighter than others in terms of energy consumption. As such, developers need to choose the lightest and most appropriate concepts in a power conscious software development when there is an alternative.

#### 3.3 Functional Programming and its Concepts

Some notable research works have been undertaken on Functional Programming and its concepts in terms of resource utilisation. Albert et al. [60] presented a framework to



define the upper boundary of the application's memory requirements so that the execution would not exceed the predefined memory limit. Their work was targeted at garbage collected languages which use mutable data, such as the Imperative and Object-Oriented programming languages. Meanwhile, Simoes et al. [61] have introduced a dynamic memory allocation framework that is able to determine the upper bounds of the memory requirements for lazily-evaluated higher-order functional programs.

Another resource utilisation approach was presented by Antoy & Jost [62], which is known as the target implementation design that aims at functional logic programs, such as Curry. These researchers implemented this design on a prototype known as Sprite that can perform functional computation in an effective manner. They, subsequently, compared the execution of this prototype with the Glasgow Haskell Compiler (GHC), and the results have indicated that this design is comparatively more efficient than GHC in terms of memory management and execution time.

Minutolo et al. [63] have proposed a lazy-evaluated pattern-matching algorithm to handle computational resources. They tested the algorithm in mobile devices, and the result proved that it is able to shorten the response time. This implies that the idea of applying Functional programming concepts to utilise resources is feasible, and the application can be further investigated by widening the scope to other available concepts so that more functional concepts will be useful for saving resources.

#### 3.4 Event-Driven with Graphical User Interface (GUI) Programming

The GUIs that are used in any software design, currently, emerged as a result of the field of HCI (Human Computer Interaction) and ID (Interaction Design) going through incremental and revolutionary changes coming from a myriad of different disciplines [64]. The user interfaces designed and employed in the major smartphone operating systems of today follow certain official guidelines developed by their vendors. These guidelines [65], [66], & [67] are a result of the HCI journey that has taken several years. But, the road of this evolution has always left performance considerations out of the interface development. Hence, these guidelines do not consider the design time performance consideration at the GUI level; although, this background has broadly described HCI, UX (User Experience), ID, and so forth.

The graphical user interface (GUI) of a piece of software is composed of a set of GUI elements, such as buttons, text boxes, etc., arranged in a meaningful manner that allows the user to interact with an electronic device [68]. Different elements of a graphical user interface exist on different software development platforms, including on the modern-day smartphone platforms, such as Andriod, iOS, and Windows Phone. However, there is also a large overlap amongst the GUI elements that is common to all mobile operating systems (and even desktop platforms) [69]. For example, the button is a common GUI element available in all smartphone operating systems and graphical desktop operating systems.

On a given UI, some of these elements, generally or under certain circumstances, can be replaced with other similar elements. For example, the day of the week can be



selected using a dropdown menu or radio buttons. This means that these GUI elements can be arranged in a myriad of ways to accomplish the same task, albeit some arrangements are more meaningful and easier to understand [70]. That is why user design guidelines have existed since the late 1980s, and are frequently used in software development to maximise the ease of use of a given GUI [71]. However, for a given screen of a software, the developer can come up with several arrangements of the GUI components, and all of those arrangements can completely abide by the usability guidelines. These screens can even make use of different GUI components from each other [72].

Given such circumstances, the selection of one meaningful arrangement of GUI out of many others has always either been irrational or just random. But, what if certain GUI elements, say a text box or label, etc., are faster than some other GUI elements, like a dropdown list? With such information at hand, a developer can make a more informed decision on which GUI elements can be swapped for others, when usability is not compromised, to make gains on performance. This is what this research aims to do.

It aims to create certain comparisons amongst the GUI elements that can be referred to by user interface developers or application developers concerned with system resources. In that, this research aims to open up the pathway for further explorations in incorporating performance factors in user interface designs.

# 4. An Evaluation of some key research works

Research in Programming Paradigms can be divided into four categories: They are the study of Programming Paradigm concepts in terms of efficiency (resource consumption), comparative analysis of the Programming Paradigms with the other Programming Paradigms, Evaluation of Programming Paradigms in view of certain types of software development (application domains), and the study of Multi Paradigm Programming. In the following sections, relevant literature to these four research categories is reviewed.

#### 4.1 Basic inherent issues in Programming Paradigms

All programming paradigms have their own merits and demerits. In general, we can classify the Programming Paradigm issues into two major categories. The first category is the inappropriateness of the programming paradigms for certain types of software developments; in section 4.2 the relevant literature is reviewed. And, the second category is the resource consumption overhead; in section 4.3 the relevant literature is reviewed. As a solution for these inherent faults, researchers have recommended alternative programming paradigms and have also come up with new programming paradigms [73, 74, 75, 76, 77, 78, 79, 80, 81, & 82]. Apart from these solutions, the researchers can suggest ways to use the currently popular programming paradigm concepts, effectively, as these concepts are widely accepted by the software industry to produce the software, magnificently.



# 4.2 Evaluating programming paradigms in terms of an application domain

In this section, the research studies which have evaluated programming paradigms in terms of a particular application domain are explored. No programming paradigm is suitable for all types of application development, but many of the current applications have not been developed using the appropriate programming paradigms. Hence, those applications are not reliable and are facing performance issues and other overhead. Software Engineering Researchers have confirmed, the inability of programming paradigms to be used for certain types of development with their research results, some of those research studies are discussed briefly in this section.

Kim et al. [83] has stated that ubiquitous environments manage various types of data and dynamic changes of situations in the real world. Traditional programming paradigms have a lack of straightforward features for such management and are not suitable for ubiquitous applications. Wei et al. [84] has described that traditional programming paradigms can only support the design time environment, but for ubiquitous computing, it needs to support the mobility and dynamically changing environment; so, the programming paradigms should support both the design and running time environments. The traditional programming paradigms cannot adapt to the adversity, complexity, dynamics, and levity of ubiquitous computing environments.

The Object-Oriented paradigm also has many shortcomings when applied to pervasive systems [85]. Weis, Braker, & Brandle [85] have mentioned that a set of objects are deeply involved with each other because an object can hold the references of many other objects, and they also perform synchronous method invocation on each other. Due to this dependency, if an object vanishes for some reason, such as weak connectivity or power off, the whole system's behaviour will get unpredictable. Security related issues also arise for this kind of dependency in the Object-Oriented paradigm. As pervasive applications depend on changing environments, so it is very hard for the Object-Oriented program developers to automatically adopt the pervasive applications and maintain security mechanisms with the changing environment. In pervasive applications, devices are small and resource restricted, so traditional programming paradigms are not suitable for the efficiency of mobile devices in terms of bandwidth, memory, CPU usage, and resource optimisation.

Mattos, C.B. and Carro, L. [86] have studied the memory and power consumption of OOP in embedded system applications. Whilst developing embedded applications, performance, power, memory consumption, and other requirements must be considered. OOP tends to cause overhead in terms of these requirements and, unfortunately, the software industry has had to accept it due to the other benefits.

In general, researchers are targeting Object-Oriented programming paradigms in terms of various application domains. This might be due to its dominance in the software industry market. Most of the successfully running Object-Oriented software systems are large enterprise-level systems. Some of the other programming paradigms are significantly contributing to the software industry, but are mostly tied with the Object-



Oriented programming paradigms in multi-paradigm software systems. Therefore, the other paradigms are also the reason for this incompatibility issue. Hence, research in this area should cover all of the programming paradigms which are involved in designing a software solution that would improve the accuracy of the research results. The benefits of this multi-paradigm approach are briefly stated in Section 4.4.

#### 4.3 Comparative analysis of the Programming Paradigms

In the Programming Paradigms area, some of the researchers comparatively analyse the efficiency of mainstream programming paradigms with each other. Generally, researchers have evaluated the Object-Oriented programming against other mainstream programming paradigms. In some of the studies, performance was measured in terms of processor time or CPU time, and in other research performance was measured in terms of memory usage or energy consumption.

Apart from this, some other researchers measured the efficiency of the programming paradigms based on how easy it was to implement a particular system using two different programming paradigms.

Dingle, A. [87] has stated that, Object-Oriented programming is inefficient compared to Procedural (Imperative) programming. The overhead caused by OOP is the major factor which is unacceptable in embedded programming. Alexander, C. [88] aimed to evaluate the effect of OOP in comparison to the traditional Procedural programming style, on both power and performance in embedded processors. The research concluded that the memory consumption for both programming styles where almost the same, but OOP consumed more power and as such, OOP may not be suitable for adoption in power critical environments, such as embedded systems.

In line with the research of Alexander, C. [88], Sumil, D.; Jamwal, S. S.; and Devanand [89] also compared OOP against Procedural programming in embedded systems, and the research concluded that the speed of execution in Procedural programming is higher than in OOP by 8%.

Other related studies, such as Barnes and Hopkins [90], Da Penha et al. [91], Chatzigeorgiou [92], and Harrison *et al.* [93] have discussed the relationship between the programming paradigm and resource utilisation. Barnes and Hopkins [90] found that, Object-Oriented implementation took more memory and processor time compared to the implementation with the Imperative (structured) programming paradigm, but the Object-Oriented programming paradigm was easier to program and implement the system compared to implementing the system with the Imperative (Structured) programming paradigm.

Besides that, studies by Shin and Cury [94] and Ahearn *et al.* [95] cited in Barnes and Hopkins [90] also support this argument by claiming that, with the adoption of the Object-Oriented approach, the ease of modelling was achieved where pragmatic implementations were considered.

An experiment was conducted by Da Penha *et al.* [91] to investigate the performance of programming paradigms and languages using multi-threading on digital



image processing. The comparisons amongst the programming paradigms, such as the Object-Oriented and Procedural programming paradigms, were conducted with sequential and parallel image convolution implementation. Java and C++ programming languages were utilised to develop the programs. They found that the Object-Oriented programming paradigm took more time to respond than the Imperative (Procedural) programming paradigm.

According to Chatzigeorgiou [92], the performance of the Procedural programming paradigm was better as compared to the Object-Oriented programming paradigm, and Harrison et al. [93] have found that Object-Oriented yielded faster execution time and compilation time, and was easier to implement as compared to the Functional programming paradigm. The application domain used in this experiment was an image analysis algorithm. However, in a similar research by Pankratius et al. [26], they refuted the claim that the Functional style was bad in performance. The average run time of Scala was proved to be comparable to Java, and for the run-times with smaller values, Scala had actually performed better than in Java. At this point in time, these kinds of viewpoints may not be appropriate as the current development scenario is based on multi-paradigm programming languages. Pankratius et al.'s [96] research result against Chatzigeorgiou [92] was one of the valid supports for this argument.

#### 4.4 Multi-Programming Paradigms

Contemporary software developments and development platforms support multiparadigms. Basically, all mainstream programming languages support multi-paradigm development, although each language design is based on a specific programming paradigm as a core. For example, Java is an Object-Oriented Programming Language but it supports Imperative, Event-Driven with GUI programming, Functional programming and others as specified in Table 1.

Multi-paradigm programming, allows the programmer to design a system with a number of different paradigm principles. The efficiency of multi-paradigm based software system solutions is always better than the single-paradigm software systems. The use of multi-paradigm programming techniques, could lower implementation costs, and result in more reliable and efficient applications [97].

All the current software project designs use the programming concepts from various programming paradigms. In Fig. 3, a sample scenario is depicted. Likewise, all current operating systems support multi-paradigms by providing a wide set of tools and frameworks. With that, any research on programming paradigms should cover all of the mainstream programming paradigms which are used in the software design. However, previous research in this field had only addressed a particular programming paradigm or a programming concept [96, 98, 99, 100, 101, & 102].





Fig. 3: Three-tier architecture with the application of Programming Paradigms

Researchers in this area, mainly focus on the multi- paradigm software design and the interaction between the different programming paradigms. In the following paragraphs, the related work of a few PhD theses are briefly summarised.

Coplien [103] has demonstrated in his research work that each multi-paradigm approach to software development has to deal with the issue of selecting an appropriate paradigm for a given application domain concept, at least to some extent. According to him, this process is seen as a mapping between the application (problem) and the solution domain. In line with this research, Valentino [104] proposed a new method of multi-paradigm software development with feature modelling in his PhD thesis, which improves the paradigm selection process to model both the application and solution domain. Its output is a set of paradigm instances annotated with the information about the corresponding application domain concepts and features.

Diomidis and Spinellis [105] have proposed the use of object-oriented design techniques as a method for encapsulating programming paradigms within multiparadigm applications, and for abstracting common characteristics across paradigms. In another research, Coplien [103] has presented a broad design method called the multiparadigm design. The main aim of this research [103] was to understand how to build systems that capture the structure of their domains in ways that support intentionality, software comprehension at the system level, and greater ease of evolution.

Basically, researchers in this area disregarded the resource utilisation perspectives of the multi-paradigm software design and developments. A rigorous research on the resource consumption aspects of programming paradigms, when they collaborate with other programming paradigms to achieve programming solutions, would help to achieve resource efficient software solutions.

# 5. Conclusion

The Programming Paradigm domain is the kernel in any software design. The Programming Languages and any software design cannot be developed without the role of Programming Paradigms. According to Campbell-Kelly [106], software possesses a great opportunity for innovation and increased efficiency. As such, further research in programming paradigms can be deemed a good way to mine this opportunity. Basically,



not much effort is being given to making software more efficient as the developers mainly concentrate on the achievement of the required solution.

Based on the above literature's knowledge at hand, the following thoughts can be deduced, whereby achieving the objectives of this paper. As per the objectives of this paper, the following content provides the concluding remarks for the academic and research community in this domain.

As for the computing education at this point of time, principally, the mainstream programming paradigms, such as Imperative (Structural and or Procedural), Object-Oriented, Functional, and Event-Driven with GUI programming, should be covered in the curriculum. Additionally, Logic Programming and Concurrent Programming could be covered as they have a notable application base in the current development scenario.

In another article, Selvakumar [107] has discussed a teaching method of the Programming Paradigms and their relevant programming languages. This article would help to teach the programming paradigms together with the respective programming languages for an application development domain. In general, the current curriculums focus on a single programming paradigm, actually, teaching multi-paradigms is appropriate for the computing education as current software designs cannot be accomplished with a single programming paradigm. For example, Imperative, Object-Oriented, and Event Driven with GUI programming can be covered in a single module and, in this case, web application development can be considered as an application domain.

In the case of Programming Languages, with the aid of one or two Programming Languages, all of these Programming Paradigm approaches can be delivered. However, in the application point of view, a Programming Language can be used for each application development, such as web application development, desktop application development, mobile application development, and the embedded application developments.

Fig. 1 and Fig. 2 can be referred to, in order to choose the appropriate programming languages to cover all of these mainstream programming paradigms and all types of application development domains. In some cases, software developers prefer different languages basically due to their interest, experience, support from the vendors, supporting tool availability, etc.; thus, the name of the programming languages are not suggested here. Therefore, in the computing curriculums, a total of four modules/subjects might be sufficient to cover the main-stream programming paradigms, their relevant programming languages and all of the four application development domains: the web, desktop, mobile, and embedded.

As for the research point of view, basically, all of the categories of research in this domain, which were described in section 4, should focus on the multi-paradigm aspects. Mostly, the previous similar research works had only addressed a particular programming paradigm or a programming concept. In the case of a research study on the efficiency of a programming design, not only should the study focus on the core programming paradigm, all other associated programming paradigms and their concepts should be considered in order to improve the accuracy of the research outcomes.



The researchers have suggested a new programming paradigm for the programming paradigm inherent problems as a solution; but instead, researchers have to identify a way to use the main-stream programming paradigms and their concepts, effectively, as most of the successfully running current projects are based on these programming paradigms.

Apart from this, researches on the resource utilisation aspects of programming paradigms, have to consider all of the other programming concepts from other programming paradigms which contribute to the software design. This would help to achieve the optimal software solutions. Finally, the researchers need to look at every possible aspect, layer, and architecture to get the software on par with the hardware in terms of efficiency.

# Acknowledgments

Thanks to God for making this paper successful. Sincere thanks to our university management for the moral support and resources provided during the research.

# References

- [1] Daniel G. Bobrow, "If Prolog is the answer, what is the question", in FGCS, Tokyo, Japan, 1984, pp. 138-145.
- [2] Linda Wieser Friedman, Comparative Programming Languages: Generalizing the Programming Function. Prentice Hall, 1991.
- [3] Pamela Zave, "A compositional approach to multiparadigm programming," IEEE Softw., vol. 6, no. 5, pp. 15-25, Sept. 1989.
- [4] Peter Van Ray and Seif Haridi, Concepts, Techniques, and Models of Computer Programming. MIT Press, 2004.
- [5] P.Van Roy and H. Seif , Concepts, Techniques, and Models of Computer Programming. MIT Press, Cambridge, MA, 2004.
- [6] P.Van Roy , The Principal Programming Paradigms, Poster version 1.08. [Online] Available from: www.info.ucl.ac.be/~pvr/paradigms.html, 2008.
- [7] Stephen Cass, [Online] Available from: http://spectrum.ieee.org/computing/ soft ware/the-2015-top-ten- programming-languages.
- [8] Stephen Cass, [Online] Available from: http://spectrum.ieee.org/computing/soft ware/the-2016-top-programming-languages
- [9] Zuhud et al., "A preliminary analysis on the shift of programming paradigms," in ICT4M, Isesco Hay Rabat, Morocco, 2013, pp. 1-5.
- [10] Souza et al., "An Experimental Study to Evaluate the Impact of the Programming Paradigm in the Testing Activity," Latin-american Center Informatics Stud. (CLEI) Electron. J., vol. 15, no. 1, pp.1-13, April 2012.
- [11] Peter Van Roy and Seif Haridi, Concepts, Techniques, and Models of Computer programing. MIT press, 2004.



- [12] Arvind Kumar Bansal, Introduction to Programming. CRC Press, 2013.
- [13] Doris Appleby, Programming Languages: Paradigm and Practice. McGraw-Hill, 1991.
- [14] David Watt, Programming Language Concepts Paradigms. Prentice Hall PTR, 1993.
- [15] B. Allen Tucker and E. Robert Noonan, Programming Languages. 2nd ed. McGraw-Hill Eductaion, 2006.
- [16] Gabbrielli et al., Programming Languages: Principles and Paradigms. New York: Springer-Verlag London, 2010.
- [17] Peter Van-Roy and Seif Haridi, Concepts, Techniques, and Models of Computer Programming. MIT Press, 2004.
- [18] Doris Appleby and J. Julius Vandekopple, Programming Languages: Paradigm and Practice. 2nd ed. New York: Mcgraw-Hill, 1997.
- [19] B. Allen Tucker and E. Robert Noonan, Programming Languages: Principles and Paradigms. 2nd ed. Boston: McGraw-Hill Education, 2007.
- [20] Quintao Pereira et al, Programming Languages. Brazil: Springer-Maceio, 2014
- [21] A. Raphael Finkel, Advanced Programming Language Design. Addison-Wesley, 1996.
- [22] W.Robert Sebesta, Concepts of Programming Languages. 7th int. ed. Addison-Wesley, 2006.
- [23] Philip Roberts. (2013, April 2). Imperative vs. Declarative. [Online]. Available :http://latentflip.com/imperative-vs-declarative/.
- [24] Microsoft Developer Network. Functional Programming vs. Imperative Programming. [Online]. Available: http://msdn.microsoft.com/enus/library/bb669144.aspx.
- [25] Zuhud et al., "A preliminary analysis on the shift of programming paradigms," in ICT4M, Isesco Hay Rabat, Morocco, 2013, pp. 1-5.
- [26] D. O. Da Penha et al, "Performance evaluation of programming paradigms and languages using multithreading on digital image processing," in Proc. 4th WSEAS Int. Conf. Appl. Math. and Comput. Sci., 2005, pp.3–8.
- [27] R. Machado, "An Introduction to Lambda Calculus and Functional Programming" in IEEE Theoretical Comput. Sci. Conf., Rio Grande, 2013, pp. 26-33.
- [28] P. Hudak, The Haskell School of Expression: Learning Functional Programming through Multimedia. Cambridge: Cambridge University Press, 2000.
- [29] M. Fenwick et al., "An Open-Source Sandbox for Increasing the Accessibility of Functional Programming to the Bioinformatics and Scientific Communities," in 9th Int. Conf. Inform. Technology – New Generations, Las Vegas, 2012, pp. 89-94.
- [30] F. Stephen, "Event-Driven Programming:Introduction, Tutorial, History," vol. II, no. 1, pp. 1-59, Aug. 2006.
- [31] M. M. ELssaedi, "Event-Driven Programming: A new approach to Event-Driven Programming," vol. I, no. I, pp. 1-157, Aug. 2008.



- [32] Dataflow Programming Concept, Languages and Applications Tiago Boldt Sousa1,2 tiago.boldt@fe.up.pt 1 INESC TEC (formerly INESC Porto) 2 Faculty of Engineering, University of Porto Campus da FEUP Rua Dr. Roberto Frias, 378 4200 - 465 Porto, Portugal.
- [33] Johan Montagnat et al., "A data-driven workflow language for grids based on array programming principles", ACM Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science, New York, USA, 2009.
- [34] Reflection in logic, functional and object-oriented programming: a Short Comparative Study. Proc. of the IJCAI'95 Workshop on Reflection and Metalevel Architectures and their Applications in AI. pp. 29–38. August, Canada, 1995.
- [35] M. H. Austern. Generic Programming and the STL Professional computing series. Addison-Wesley, 1999.
- [36]
   Francesca Rossi, Peter van Beek, Toby Walsh, "Constraint Programming".

   [Online]
   Available
   from:

https://cs.uwaterloo.ca/~vanbeek/Publications/kr\_handbook06.pdf.

- [37] David, W, "The concepts of concurrent programming", [Online] Available from ftp://ftp.sei.cmu.edu/pub/education/cm24.pdf.
- [38] X. Meng, "Designing approach analysis on small-scale software performance testing tools," in Int. Conf. Electronic and Mechanical Engineering and Information Technology, 2011 © IEEE. doi: 10.1109/EMEIT.2011.6023983.
- [39] L. Richie et al., "An object-oriented simulation–optimization interface," Comput. and Structures, vol. 81, no. 17, pp. 1689-1701.
- [40] P.Van Roy and H. Seif , Concepts, Techniques, and Models of Computer Programming. MIT Press, Cambridge, MA, 2004.
- [41] A. B. Tucker and R. E. Noonan, Programming languages Principles and paradigms. New York: McGraw-Hill, 2007.
- [42] Case Study: Oracle, "Class StringBuilder", Oracle, 2013.
- [43] Case Study: Oracle, "Class StringBuffer", Oracle, 2013.
- [44] Case Study: Oracle, "Class Vector<E>", Oracle, 2013.
- [45] Case Study: Oracle, "Class ArrayList<E>", Oracle, 2013.
- [46] M. Rubio-Sánchez, "Tail Recursive Programming by Applying Generalizatio," in Conf. Innovation and Technology Comput. Sci. Educ., 2010.
- [47] M. C. ER, Performance Evaluations of Recursive and Iterative Algorithms for the Towers of Hanoi Problem. Computing, 1986.
- [48] S. Schaeckeler and W. Shang, "Stack Size Reduction of Recursive Programs," in Inter. Conf. Compilers, Architecture and Synthesis for Embedded Syst., 2007.
- [49] Venkat Subramaniam, 2012. Scala for the Intrigued Recursions and Tail Call Optimization. Issue #35, May 2012. The Pragmatic Programmers.
- [50] Y. A. Liu and S. D. Stoller, "From Recursion to Iteration: What are the Optimizations?," in Proc. ACM workshop on Partial Evaluation and Semantics-Based Program Manipulation, 2000.
- [51] S. Schaeckeler and W. Shang, "Stack Size Reduction of Recursive Programs," in Inter. Conf. Compilers, Architecture and Synthesis for Embedded Syst., 2007.



- [52] P. Chandran, "Updating and Extending the Open Source Java Coverage Tool, Coverlipse," M.S. thesis, Uni. Edinburgh, 2009.
- [53] C. B. Mattos and L. Carro, "Making Object Oriented Efficient for Embedded System Applications," ACM Integrated Circuits and Systems Design, Florianopolis, 2005, pp. 104-109.
- [54] M. Boasson, "Embedded Systems Unsuitable for Object Orientation," Reliable Software Technologies, vol. 2361, no. 1, 2002, pp. 1-12.
- [55] S. W. Haney, "Is C++ Fast Enough for Scientific Computing?," Comput. Physics, vol. 8, 1994, pp. 690-694.
- [56] A. D. Robinson, "C++ Gets Faster for Scientific Computing," Comput. Physics, vol. 10, 1996, pp. 458-462.
- [57] B. Calder et al., "Quantifying Behavioral Differences Between C and C++ Programs," J. Programming Languages, vol. 2, 1994, pp. 313-351.
- [58] A. D. Robinson, "The abstraction penalty for small objects in C++," in Parallel Object-Oriented Methods and Applications Conf., Santa Fe, New Mexico, 1996.
- [59] Kayun, C., Chonawat, S., (2007a). Object-Oriented Programming Strategies in C# for Power Conscious System. World Academy of Science, Engineering and Technology, 10 (10), p587-592.
- [60] E. Albert et al. (2010). Parametric Inference of Memory Requirements for Garbage Collected Languages. [Online] Available: http://dl.acm.org/citation.cfm?id=1806671.
- [61] H. Simoes et al., "Automatic Amortised Analysis of Dynamic Memory Allocation for Lazy Functional Programs," in Proc. 17th ACM Intern. Conf. Functional programming, Copenhagen, NY, 2012, pp. 165-176.
- [62] S. Antoy and A. Jost. (2013). A Target Implementation for High-Performance Functional Programs [Online]. Available: http://web.cecs.pdx.edu/~antoy/homepage/publications/tfp13/paper.pdf.
- [63] A. Minutolo et al., "A Lazy Evaluation Approach for Mobile Reasoning in DSSs," in 12th IEEE Intern. Symp. Computational Intell. and Inform., Budapest, 2011.
- [64] J. Grudin, A Moving Target The Evolution of Human-Computer Interaction. 2012.
- [65] Microsoft Developers Network. (2015). Introduction to Universal Windows Platform (UWP) apps for designers - Windows app development.
- [66] Google design guidelines. (2015). Introduction Material design Google design guidelines. [online] Available: http://www.google.com/design/spec/material-design/introduction.html.
- [67] Apple iOS HCI Guidelines. (2015). iOS Human Interface Guidelines: Designing for iOS.
- [68] C. Ghaoui, Encyclopedia of human computer interaction. Hershey PA: Idea Group, 2006.
- [69] X. Meng, "Designing approach analysis on small-scale software performance testing tools," 2011 Intern. Conf. Electronic & Mechanical Eng. and Inform. Technology.





- [70] J. Grudin, "Three Faces of Human-Computer Interaction," IEEE Annals Hist. Comput., vol. 27, no. 4, pp.46-62, 2005.
- [71] D. Norman, The psychology of everyday things. New York: Basic Books, 1988.
- [72] Thompson and John, Developer's Guide to UI. Oracle, 2012, pp.16-18.
- [73] Torben et al., *Towards a Programming Paradigm for Pervasive Applications based on the Ambient Calculus*. Microsoft Research, Cambridge, 2011.
- [74] Mbayashi, Y.; Ledgard, H.F.; Dept. of Comput. & Inf. Eng., Nippon Inst. of Technol., Saitama, Japan, An experiment on a new programming paradigm, IEEE, 2005.
- [75] Jan Vitek, New Paradigms for Distributed Programming Jan Vitek ,Object Systems Group, University of Geneva, Geneva, Switzerland, 1997.
- [76] Dennis Gannon, *Programming Paradigms for Scientific Problem Solving Environments*. Woco9, Prescott, 2006.
- [77] L. Ling Shang, "Experiments on Programming paradigms for large scale scientific computing, grids, desktop grids and private clouds," M.S.thesis, 2010.
- [78] Manuel Oriol, *The Internet Programming Paradigm*. CiteSeer 6M, 2011.
- [79] M.Y.Kim et al., *A New Programming Language for Ubiquitous Applications*. 2013.
- [80] S. A. Salustri and R. D. Venter, *A new programming paradigm for engineering design, engineering with Computers.* Springer Verlag, London limited, 1994.
- [81] Johan Glimming, Thorsten Altenkirch and Patrik Jansson, What is the next Programming Paradigm?, Second International Software Technology Exchange Workshop, Kista, Sweden, 2012.
- [82] H. Jin et al., "Testing new programming paradigms with NAS Parallel Benchmarks," NASA Technical Report Server, 2000 © NASA Ames Research Center; Moffett Field, CA United States. doi: 20000064623
- [83] M. Y. Kim et al. A New Programming Language for Ubiquitous Applications [online]. Available: http://www.ics.uci.edu/~lopes/bspc05/papers/kim.pdf.
- [84] W. Wei et al., "Allotropy Programming Paradigm for Ubiquitous Computing Environment," in Int. Conf. Convergence Inform. Technology, pp. 514-521, 2007.
- [85] T. Weis et al. (2013) Towards a Programming Paradigm for Pervasive Applications based on the Ambient Calculus [online]. Available: http://wenku.baidu.com/view/f0cfddc72cc58bd63186bdac.html.
- [86] C. B. Mattos and L. Carro, "Making Object Oriented Efficient for Embedded System Applications," ACM Integrated Circuits and Syst. Design, pp. 104-109, 2005.
- [87] A. Dingle, Improving C++ Performance in Temporaries. Seattle Univ., WA, USA, 1998.
- [88] C. Alexander, "Performance and power evaluation of C++ object-oriented programming in embedded processors," Inform. and Softw. Technology, vol. 45, no. 4, pp. 195-201, 2003.
- [89] Sunil Dutt et al., "Object Oriented Vs Procedural Programming Embedded Systems," Intern. J. Comput. Sci. & Commun., vol. 1, no. 2, pp. 47-50, 2010.
- [90] Barnes and Hopkins, The impact of programming paradigms on the efficiency of an individual-based simulation model. UK, science direct, 2001.



- [91] D. O. Da Penha et al., "Performance evaluation of programming paradigms and languages using multithreading on digital image processing," in Proc. 4th WSEAS Intern. Conf. Applied Math. and Comput. Sci., pp.3–8, 2005.
- [92] A. Chatzigeorgiou, "Performance and power evaluation of C++ object-oriented programming in embedded processors," Inform. and Softw. Technology, vol. 45, no. 4, pp.195-201, 2003.
- [93] R. Harrison et al., "Comparing Programming Paradigms: an Evaluation of Functional and Object-Oriented Programs," Softw. Eng. J., vol. 11, no. 4, pp. 247-254, 1996.
- [94] Y. J. Shin and P. Cury, "Exploring community dynamics through size dependent trophi interactions using a spatialized individual-based model," Aquati Living Resources, vol. 14, no. 2, 2001.
- [95] S. C. Ahearn et al., "TIGMOD: an individual-based spatially explicit model for simulating tiger/human interaction in multiple use forests," Ecological Modelling, vol. 140, no. 81, 2001.
- [96] V. Pankratius et al., "Combining Functional and Imperative Programming for Multicore Software: An Empirical Study Evaluating Scala and Java," in 34th Int. Conf. Softw. Engineering, New York, 2012, pp. 123-133.
- [97] D. Diomidis Spinellis, Programming paradigm as Object classes: A structuring mechanism for Multi Paradigm Programming. 1993.
- [98] NELSON, H. J., MONARCHI, D. E. & NELSON, K. M. Year. Evaluating emerging programming paradigms: an artifact-oriented approach. In: System Sciences, 1998. Proceedings of the Thirty-First Hawaii International Conference on, 6-9 Jan 1998 1998. 446-454 vol.6.
- [99] S. Antoy and A. Jost, A Target Implementation for High-Performance Functional Programs. 2013.
- [100] A. Minutolo et al., "A Lazy Evaluation Approach for Mobile Reasoning in DSSs," in 12th IEEE Int. Symp. Computational Intell. and Informatics, 2011.
- [101] Sunil Dutt et al., "Object Oriented Vs Procedural Programming Embedded Systems," Int. J. Comput. Sci. & Commun., vol. 1, no. 2, pp. 47-50, 2010.
- [102] W. D. Clinger, "Proper Tail Recursion and Space Efficiency," in Proc. ACM Conf. Programming Language Design and Implementation, pp. 174 – 185, 1998.
- [103] O. James Coplien, Multi-Paradigm Design for C++. Addison-Wesley, 1999.
- [104] Valentino Vranic, Multi Paradigm design with Feature modelling. 2004.
- [105] D. Diomidis Spinellis, Programming paradigm as Object classes: A structuring mechanism for Multi Paradigm Programming. 1993.
- [106] M. Campbell-Kelly, "The History of the History of Software", IEEE Annals Hist. Comput., vol. 29, no. 99, pp. x4, 2007.
- [107] S. Selvakumar," Teaching Programming Subjects with Emphasis on Programming Paradigms", in icaet-14 © Atlantis Press. doi:10.2991/icaet-14.2014.22.



# A Review of Data Analytical Approaches in the Insurance Industry

Noorhannah Boodhun Faculty of Computing, Engineering & Technology Asia Pacific University of Technology and Innovation 57000 Kuala Lumpur, Malaysia noorhannahb@gmail.com

*Abstract* - Insurance companies are facing a growth in their transactional data. Valuable information can be gained from the data collected by using analytical approaches. The key areas in which data analytics can be useful to insurance firms is customer level analytics, risk assessment and prediction as well as fraud detection. Commonly used techniques to identify patterns in the data sets are clustering and classification to be able to predict future occurrences of events. The main purpose of implementing analytics among the insurance firms is to better understand their customers, minimize their losses and gain a competitive advantage in the market. A review of several algorithms used to analyse insurance data is provided in this paper together with some evaluations of the different approaches.

Index Terms - Data analytics; insurance industry; risk prediction; fraud detection

# 1. Introduction

In the global era, demand of insurance is increasing and therefore insurance companies L have a large amount of data at their disposal. The insurance sector generally comprises of life, health and non-life insurance. Insurance can be defined as the hedging of the risk of an uncertain loss (Robson, 2015). Insurance companies are expanding all over the world and like other industries, they also must bear the cost of expansion. The insurance market faces threats like fraud and abuse cases as well as it encounters the risk of customer attrition (Sithic & Balasubramanian, 2013; Mohammadi, Albadvi & Teymorpur, 2014). Therefore, it is critical for the insurance companies to come up with intelligent solutions to sustain their businesses. A good approach is to implement analytics to their rich amount of data to enable better decision making. Analytics provide companies with the abilities to gain useful insights from their customer database and hence can allow the insurance companies to gain competitive advantage over their adversaries (Umamaheswari & Janakiraman, 2014). Nowadays, more businesses are shifting towards the use of data analytical techniques to improve operations (Sivarajah et al., 2017). With the revolution of technology, insurance industries have experienced many changes in the way they operate. Technological innovations, like data warehousing has allowed them



to cut down their costs of storing and accessing data. However, the conventional techniques of analysing data may not be reliable in today's world, where the existence of big data is prominent. Hence, it is essential for the insurance sector to benefit from the availability of modern data mining techniques and machine learning algorithms to scrutinize their available data.

The several data analytical approaches in different insurance industries is studied in this paper. The importance and efficacity of the techniques are elucidated and the objective of this research is to provide a review of the various methods used to analyse a broad variety of insurance data as an endeavour to improve the industry. The scope focuses on three key features, specifically, customer analytics, risk prediction and fraud detection in the insurance sector. The paper is divided as follows. Section 2 describes the application of data analytics in the insurance industry, followed by Section 3 which highlights the discussion of the literature whereby prominent limitations in the approaches for data analytics in the insurance industry are elaborated. Finally, Section 4 outlines the conclusion.

# 2. Data Analytical Approach in the Insurance Sector

In this section, the different data analytical approaches in the insurance sector will be discussed. The areas discussed will be customer analytics, risk prediction and fraud detection.

# 2.1 Customer Analytics

Customer analytics refers to studying the behaviour of customers in order to anticipate what they want from the company. It is usually carried out by using data mining techniques, specifically, predictive analytics with the objective to increase customer loyalty and enhance customer retention (Zhou & Chen, 2014). Data mining is a powerful tool allowing insurance companies to concentrate on valuable information gathered about the behaviour of their clients. It involves classification and segmentation of customers according to various criteria in order to model their behaviour and purchase patterns. As such, insurance firms can profile their customers based on their data and eventually target them with appropriate policies which they might be keen to buy (Desik et al., 2016). Insurance firms are highly dependent on customers, thus doing analytics enhances the marketing strategies of the firms and aids in cost reduction (Ravasan & Mansouri, 2015). Customer level analytics offers the potential of accurately identifying the specific policies to sell to particular customers. Researchers propose various algorithms, such as classification, clustering, association, regression, and several others, that have been widely studied and implemented on insurance data to create prediction models (Rahman et al., 2017; Hanafizadeh & Paydar, 2013).

Golmah (2014) carried out an analysis on customer segmentation in the automobile insurance industry. A case study of 23 automobile insurance companies in



Iran was used to analyse customers' choices and personalize marketing campaigns accordingly. Data mining tool used was Self Organizing Maps (SOMs), which applied unsupervised classification of the insurance customers to understand their behaviour in order to increase customer loyalty. Data collection was done by using questionnaire and the target respondents were automobile insurance customers. The questionnaire comprised of four sections collecting demographics data, automobile characteristics data, automobile insurance information and factors impacting insurance selection. Out of the 2000 questionnaires distributed via email, 1260 effective responses were used for the analysis after data cleaning. The SOM technique was implemented using MATLAB software to cluster customers into distinct groups. The prediction accuracy of the SOM was assessed using statistical measures like mean absolute error (MAE) and root mean square error (RMSE). The calculated MAE and RMSE values were 0.04 and 0.06 respectively, demonstrating that the model was precise. Moreover, a U-matrix map was used to visualize distinct clusters in the data set. The results established four clusters of customers with a set of characteristics. The clusters were those to whom cost of policies matter, quality of service matter, style of payment is crucial and those who chose their insurance policies in regard to their family and friends. The SOM method of clustering customers proved to be useful on one data set and can therefore be applied to other insurance data sets to improve marketing strategies.

Another research dealing with the segmentation of customers for the analysis of behaviour is done by Jandaghi & Moradpour (2015). They used the data mining technique, fuzzy clustering on life insurance customers. Secondary data was collected from the database of Pasargad life insurance company for a period of April to October 2014 and consisted of 1071 customers. Age, gender, number of children, job, relationship between insured and insurer, insurance term, payment method, premium, number of supplementary coverage were among the efficient attributes used for the analysis, after interviewing experts for the selection of right attributes. R statistical software was employed to carry out the data mining task of fuzzy clustering method. Findings suggested an optimal number of two clusters, which were labelled 'investment policy' and 'life security'. Customer segmentation is concluded to be an appropriate technique to improve sales and performance of the insurance company. However, a prominent limitation of this research is that the authors did not justify the accuracy of their model by any statistical means of evaluating performance of a method.

Roodpishi & Nashtaei (2015) conducted a research using insurance data in order to perform market basket analysis to better understand customer demand. Data from a car insurance company in Anzali city of Iran was used to perform the analysis. The data set was for the year 2014 and was collected from 300 insurance clients using questionnaires. The primary technique employed to mine the data was the K-means clustering algorithm which clustered the data into five optimal clusters based on the demographic attributes present in the data set, namely gender, age, occupation, education level, marital status, place of residence and income. The data mining software SPSS Clementine was the tool utilised to perform the clustering method. Further in the research, association rules were performed on each cluster using causal extraction



algorithms to identify hidden patterns in the data set and to predict customer behaviour. Moreover, the performance evaluation criteria used to validate the modelling was the sum square error (SSE). A useful implementation of the study is to target potential clients to sell appropriate policies. However, the data set used was limited to only one region which is why the findings cannot be generalised for the whole country.

Fang, Jiang & Song (2016) identified critical factors to predict customer profitability in the insurance industry using big data analytical approach. Data was obtained from a commercial industry company in Taiwan. Customer profitability is the segmentation factor to predict valuable and nonvaluable clients. The data set consisted of 25,000 observations with several attributes such as premium paid, claim amount, payment type, guarantee years, age, gender, occupation, region. The research aimed to develop a new model for the prediction of customer profitability. Liability reserve was the new variable added to distinguish the model from existing traditional customer profitability models, which only deal with premium and claim amounts. The new model was tested using 7:3 ratio of training and testing data set. Random forest regression was applied to forecast the insurance customer profitability. Moreover, the model was compared against linear regression model, decision tree, support vector machine (SVM) and generalised boosted model. Random forest regression provided higher accuracy compared to the other models, accounting for 99.03% of variation in the model. However, a pertinent drawback in the model was that it did not include customer income data, which could affect the accuracy of the model.

A study conducted by Delafrooz & Farzanfar (2016) employed clustering method in the insurance industry to determine the lifetime value of customers. Customer lifetime value is another way of assessing the behaviours of customers by segmenting them and therefore evaluating their profitability for the company in order to improve customer relationship management systems. The customer lifetime value is calculated using metrics like learning cost, customer loss rate, discount rate, costs of maintenance, periodic income, period of time and profit margin. The customers are segmented according to the percentage of customer lifetime value, most precisely a benefit segmentation is carried out. Data used in this research was obtained from the companies' database and comprised of 10 types of insurance companies whose customers have been studied for a period of four years. Based on the customer lifetime value calculations, third party insurance customers had the highest amount of customer lifetime compared to freight insurance which showed the lowest amount. Results show that the customers were segmented into classes, namely 'gold', 'silver', 'bronze' and 'tin', with gold being those with the highest profitability.

Ansari & Riasi (2016) used neural networks and linear regression analysis to identify factors impacting the customer loyalty among start-up insurance companies. The approach was to evaluate the performance of both techniques in analysing customer loyalty. A total of 389 customer data from 10 start-up insurance companies in Iran were collected using survey technique. Three models were developed for the linear regression analysis using target variables perceived value, customer satisfaction and customer loyalty for each. Besides, artificial neural networks were created using MATLAB software



and several possible networks were tested to find the best suited one for the data set. Both techniques suggested that perceived value and customer satisfaction were significant predictors of customer loyalty. However, findings obtained from the artificial neural network models had lower error rates compared to the regression models. Hence, it could be concluded that artificial neural networks is best suited for analysing customer loyalty in the insurance market.

#### 2.2 Risk Prediction

Insurance companies highly practice risk management because insuring a policyholder means that the risks are transferred from the customer to the firm. Insurance firms profile and assess the risks of customers to deduce the amount of premium that must be paid (Polyakova, 2016). Traditionally, risk is calculated using actuarial formulas, yet now, with the presence of data analytics more research is focusing on the deduction of new formulas with more accurate factors impacting the risk amount of a customer. Nonetheless, risk prediction based on business risk can also be explained as the risks involved with customer churn, which is mostly the focus of this paper. The literature present on customer churn analysis is considerable. Customer churn analysis is a main concern among numerous industries, for instance the retail industry, telco industry and banking institutions (Almana, Aksoy & Alzahran, 2014; Vafeiadis et al., 2015; Amoo et al., 2015). Some extensively used techniques to evaluate the risks involved in the insurance companies are decision trees, neural networks, Naïve Bayes and regression analysis. The predictive analysis of whether a customer is likely to churn is fundamental for any company so as it can ameliorate its customer relationship management to deter the customers from terminating their relationship with the company (Rodriguez & Shin, 2013).

Günther et al. (2014) performed a study in which they created a statistical model for the prediction of customer churn. They obtained data from Gjensidige, which is Norway's largest non-life insurance company. The data involves health, home and car insurance policy holders. The data is of time series nature and is for a 19 months' period, which dates from November 2007 to May 2009. The data set contains numerous attributes, specifically, yearly premium, age of customer, gender, partner, discount programme, car insurance policy, home insurance policy, number of home insurance policies, health insurance policy and customer lifetime. A random sampling techniques was employed to select 160,000 customers from the database, which eventually decreased to an effective sample size of 127,961 customer data after data cleaning and reduction. Furthermore, a generalised additive model (GAM) approach was used to identify linear relationships between the explanatory attributes and some attributes were log transformed to meet linearity conditions. The model was based on the linear regression model and the analysis identified significant attributes which were used in building the prediction model. The receiver operating characteristic (ROC) curve was utilised to evaluate the prediction performance. Moreover, the regression model was validated using a new data set for the period June 2009 to January 2010 and the



prediction performance was tested for each month. Results elucidated that the model was accurate.

Likewise, Goonetilleke & Caldera (2013) conducted a research on customer churn prediction. However, the study was designated to life insurance domain, using data from a life insurance company in Sri Lanka for policies starting from the year 2002 to 2003. The authors used classification techniques, namely decision trees, neural networks and logistic regression method to analyse the data. Initially four class labels were established for the policies, explicitly, 'lapse', 'open', 'paid-up' and 'mature' to determine the status of the policies in the data set. The approach involved selection of attributes with the help of domain experts followed by initial exploration of the data set to find any significant patterns in visualizations like stacked bar charts and by statistical tests like chi-squared testing. The set of attributes dealt with were demographic information, policy details and number of other policies the customer has with the company. After initial data cleaning and pre-processing, 21 variables were selected. Yet, other techniques, namely, correlation-based feature selection subset evaluation, gain ratio and information gain were employed to rank the most importance attributes and eventually only 8 most important were used for the data mining task. A prediction model was built using the Weka package software by applying decision trees, neural networks and logistic regression. The different models were evaluated using the performance criteria, prediction accuracy, ROC curve and AUC value. Besides, cost sensitive learning strategies were employed to the data set to take care of skewness in the class distribution. Upon application of the strategies, findings suggested a 90% accuracy in capturing customer attrition by the prediction model.

Rodpysh (2013) came up with a formula to calculate the customer churn index using insurance data and applied data mining techniques to evaluate the formula. The objective of this investigation was to propose a formula to calculate the optimal customer churn in order to be able to assess risky situations. Data used in this study originated from a third-party insurance database in Guilan province for the period July to September 2011. The proposed churn formula comprised of six variables, namely, satisfaction from competing companies, overall satisfaction of company's products, probability of discontinuation of services in the future, likelihood of using competing company's products, likelihood of using the company's products and service and company recommendations. It was explained that customers with a calculated index of 0.5 and above have a high probability of churn. The formulation was compared to five other indices of churn, namely, 'may terminate the contract and service', 'extension of contract and services', 'advised the company to others', 'satisfaction of rival companies' and 'overall satisfaction of the company' using six different classification methods to determine customer churn and the accuracy of each model was assessed. The six classification models employed were QUEST decision tree, C5.0 decision tree, CHAID decision tree, CART decision tree, Bayesian Network and Neural Network. The models were evaluated using the methods overall accuracy, profit function, ROC curve and Lift Index. Findings suggested a better accuracy in predicting customer churn using the proposed formula instead of individual indicators.



Kašćelan, Kašćelan & Novović Burić (2016) used a nonparametric data mining approach for risk prediction in the car insurance industry. Data used was collected from third party motor insurance company in Sava Montenegro. The data analysed was for the period 2009 to 2011 and consisted of 35,521 policies data with attributes such as region, age, sex, type of vehicle, number of claims per policy, years of policy ownership and mean claim size. The research aimed to demonstrate the effectiveness of using data mining techniques to predict risk instead of the standard techniques used for risk assessment. K-means clustering method was applied to form clusters of similar data points and 12 clusters were obtained. Support vector regression (SVR) was performed on the clusters with average claim cost being the target variable. SVR was used to estimate claim size and the performance evaluation statistics showed that the technique had relative error of less than 10% among most clusters. Logistic regression was applied using number of claims as the target variable, to estimate the probability of a claim occurring. Logistic regression model had an accuracy of around 80%. The proposed approach revealed that data mining techniques are fairly accurate when predicting risk in the car insurance industry, even with small data sets. However, the techniques were not tested against other data mining algorithms which could have provided better results.

Similarly, Rose (2016) used a nonparametric approach to analyse insurance data. However, the framework proposed was machine learning algorithms. The aim of the research was to conduct machine learning for determining a simple predicting formula for plan payment risk adjustment and evaluate the likelihood of improving risk adjustment. Data originated from Truven MarketScan database from which a random sample of 250,000 data was chosen and was for the period 2011-2012. The data consisted of enrolment and claim data from private health plans and the attributes used for the research involved age of client, gender, region, inpatient diagnosis categories and 74 created Hierarchical Condition Category (HCC). Analysis involved methods like linear regression, penalised regression, artificial neural networks, decision trees, random forests. To assess the performance of the algorithms on a data set a cross-validation method was applied, which refers to assigning measures of performance to each algorithm to validate whether it is useful when applied to novel data. The optimal choice of prediction function was denoted as the super learner and has been developed by the super learning algorithm, which is the incorporation of important components from multiple algorithms to make up a single combined algorithm. Findings suggested that the super learner model outperformed the other algorithms studied.

#### **2.3 Fraud Detection**

Currently, insurance fraud and abuse is a growing concern among the insurance industries. Fraud refers to intentional misrepresentation of certain information to make false claims. Insurance fraud is detrimental to the insurance firms as they incur great losses when false claims are made. The fraudulent cases can be found among any type of insurance, yet it shows more occurrence in the healthcare insurance sector since



physicians have been found falsifying medical reports for policyholders. Statistics reveal that the United States of America incurs over 30 billion dollars every year due to health insurance frauds (Rawte and Anuradha, 2015). Data analytics is providing great support in insurance fraud detection, since recently companies are facing an upsurge of transactional data that is being collected daily (Bănărescu, 2015). Over the past years, fraud detection methods have received much interest among researchers (Sithic & Balasubramanian, 2013; Muhammad, 2014; Li e al., 2016; Wang et al., 2017). Many algorithms have been applied to provide models where fraud and abuse cases can be spotted efficiently as an attempt to improve business processes and reduce unnecessary losses.

Joudaki et al. (2016) identified potential fraud and abuse from health insurance data set by using data mining approach. Data was collected from the Social Security Organisation (SSO) from Iran and consisted of general physicians' drug prescription claims data for 612,804 outpatient drug prescription claims in year 2011. The aim was to detect potential suspects among the physicians who would be involved in fraudulent activities. Indicators of fraud and abuse were developed whereby 13 indicators were found significant to be able to classify the suspects. A hierarchical clustering method was used to identify the segment of physicians who were suspects for fraudulent practices. Discriminant analysis was carried out to verify the reliability of the clusters formed. It was found that the indicators employed to perform the clustering showed satisfactory performance as when tested on a new sample data set, fraud suspects were detected with 98% accuracy and abuse suspects with 85% accuracy. The research demonstrated the effectiveness of applying data analytical approach to health insurance data and this concept can be useful to streamline auditing practices towards suspect groups instead of all physicians.

Another research on fraud and abuse detection in the health care insurance is undertaken by Kose, Gokturk & Kilic (2015) who developed an interactive system called eFAD suite to detect fraudulent activities. The framework was designed to be independent of actors who involve individuals like insured or physicians and commodities, which are the medications that the insurance company pays for. The framework is based on interactive machine learning approach which means that the system is flexible to changes and allows experts to interact with the data. Transactional data was used to identify suspicious fraudulent and abuse cases. To achieve a proactive analysis, which refers to the online detection of fraud and abuse, and enhance performance, a two-stage data warehousing method was employed. The authors used pairwise comparison method for weighting the actors and commodities and clustering of similar actors was done by expectation maximization method. The risk associated with claims and actors were determined using the z-score of the attributes in the study. Finally, a dashboard was created for visualization purpose. The visualization tool gives analysis results based on the input attributes and risk factors. The model accuracy when tested gave an overall accuracy statistic between 70.8% to 89.6%.

Thornton et al. (2014) used a different methodology to analyse fraudulent activities from healthcare insurance data, which is the outlier detection method. Outlier



detection method is an unsupervised anomaly detection method commonly used to detect frauds. The analysis was performed using the dental providers case study which comprised of data in a state Medicaid program from the United States from July 2012 to May 2013, amounting to 650,000 dental claims. Attributes that were involved are claim data, healthcare provider and patient data and were analysed using an iterative outlier detection technique. Fourteen specific fraud metrics in the dental domain, were identified based on literature, discussions with domain experts and analysed cases from the U.S Federal Bureau of Investigation. The analysis techniques used included multivariate analysis, time series analysis and box-plot analysis and the outlier detection methods applied were regression analysis, K-means cluster analysis, trend deviations and peak deviations. Oracle SQL Procedures and R language scripts were utilised to conduct the analysis. The findings suggested that 12 out of 17 (71%) suspected dental providers were subjected to formal investigation by officials for committing fraudulent practices.

Similarly, Nian et al. (2016) employed an unsupervised anomaly detection method to identify fraudulent activities in the auto insurance industry. Nonetheless, they proposed a new technique which is the ranking method for anomaly detection based on spectral analysis. The proposed method involved a ranking scheme which identified the top ranked case as being the most suspicious. The method detects anomaly in the dependence among attributes in a dataset by using similarity kernels. Spectral analysis was employed to generate anomaly ranking and assist visualizations. For this analysis, to assess the performance of the proposed method, an open access auto insurance claim data set from Angoss Knowledge Seeker Software was used with 15,420 claims data from January 1994 to December 1996. The data set comprised of over 30 attributes, both ordinal and categorical. Together with the proposed method of spectral ranking analysis, one-class support vector machine (OCSVM), local outlier factor (LOF) and supervised random forest (RF) techniques were tested on the data set to detect fraudulent cases. OCSVM was obtained from the LibSVM Library while LOF was from Ddtools Library and RF from MATLAB software. The ROC curves and AUC were used as performance assessment criteria to compare the accuracy of the models. Results suggested that the unsupervised spectral ranking analysis method provided more precision in determining fraud cases compared to the other fraud detection methods.

Another research related to automobile insurance fraud detection is done by Hargreaves & Singhania (2016). They used the same data set as the previous mentioned paper, which is from the Angoss Knowledge Seeker software comprising of 31 attributes in regard to auto insurance claim data. However, their approach in detecting fraud was different. Hypothesis testing was conducted on the attributes in the data set to identify the significant variables for determining fraud cases. The Chi square test and independent sample t-test were employed to determine any correlation between the categorical attributes and to compare means of the continuous variables respectively. 20 out of 31 variables were found significant and were used to profile fraudulent and nonfraudulent claims. A set of characteristics was identified for the fraud group based on the significant variable identification. Furthermore, a total of 20 business rules were derived



to aid the identification of future fraudulent claims. Four rules were derived based on demographic characteristics of a fraud profile, ten rules were identified to test the claim characteristics of a fraud profile, another four rules were based on the vehicle characteristics and finally four rules were based on the policy type to identify fraudulent claims.

Goleiji & Tarokh (2015) used two methods to select significant factors pertaining to automobile insurance fraudulent practices. The feature selection methods used to choose the influential attributes are genetic algorithm method and correlation method. The genetic algorithm feature selection method uses learning algorithm and is similar to Naïve Bayesian method. The selection of feature based on correlation is the Pearson test which evaluates the strength of association between attributes. The data set used in this study was obtained from an automobile insurance industry comprising of 27 attributes. According to the genetic algorithm, 14 variables were crucial while for the correlation method, 13 variables were found important. Decision tree algorithm and Naïve Bayesian techniques were applied to both set of attributes and fraudulency and non-fraudulency was considered as the target variable. Results suggested that among both feature selection methods and data mining techniques, decision tree models applied on the set of attributes chosen by the genetic algorithm demonstrated the highest accuracy (93.89%).

# 3. Discussion

# 3.1 Most Popular Approach

Based on the researches reviewed in the previous section, it can be found that for customer level analytics, the most popular approaches were clustering and classification to separate the customers into various groups in order to understand their behaviour. The reason why their behaviour is studied is because the companies want to know their customers better to offer them with the right products in order to maximize their satisfaction and to ensure their retention. For risk prediction purposes, the most common approach was again classification, specifically decision trees and neural networks. Such algorithms provide a better insight on the prediction of customer churn. It is crucial for insurance firms to identify reasons and likelihood of their customer attrition rates because it can help them strive in the competitive market. As for fraud detection, the literature suggests that researchers are coming up with novel algorithms based on the existing traditional methods like clustering, classification and linear to get more accurate prediction models to detect fraudulent cases.

# 3.2 Popular Evaluation Criteria

Most of the studies applied performance evaluation criteria to measure the accuracy of their model. Out of the literature reviewed in this paper, root mean square error (RMSE),



mean square error (MSE), receiver operating characteristic (ROC) curve and area under curve (AUC) were the most applied statistical measurements for testing the accuracy of the models used.

#### **3.3 Limitations**

A few researchers did not elaborate on their source and attributes of data set. This does not allow the analysis to be reproduced with other data mining techniques using the same data set. However, the reason why the data source is not revealed is due to privacy and data protection policies of the insurance firms. The companies deal with sensitive customer data and are restricted by laws to not disclose any personal information. Certain studies also faced hindrance in their analysis because they were not given the right to access the required data or the data provided to them were masked for privacy reasons. Moreover, few papers failed to evaluate their models based on statistical proofs and performance evaluation criteria, which makes their research not robust enough to extract valuable information. Table 1 gives an overview of the different literature reviewed in this paper and their main approaches to data analytics in the insurance sector.

Area	Research	Insurance Industry	Approach	Algorithms	Evaluation Criteria
Customer Analytics	Golmah (2014)	Automobile	Classification	Self- Organizing Maps (SOMs)	Mean absolute error (MAE), root mean square error (RMSE)
	Jandaghi & Moradpour (2015)	Life	Clustering	Fuzzy clustering	No statistical evaluation
	Roodpishi & Nashtaei (2015)	Automobile	Market basket analysis	Causal extraction algorithms, k- means clustering	Sum square error (SSE)
	Fang, Jiang & Song (2016)	Healthcare	Regression	Random forests regression	Mean square error (MSE)
	Delafrooz & Farzanfar (2016)	Life and non-life	Segmentation	Benefit clustering	No statistical evaluation
	Ansari & Riasi (2016)	Not disclosed	Classification, regression	Neural network, linear	MSE, RMSE

#### Table 1: Summary of approaches



				regression, ANOVA testing	
Risk predictio n	Goonetillek e & Caldera (2013)	Life	Classification	Decision tree, neural network, logistic regression	Receiver operating characteristic (ROC) curve, area under curve (AUC)
	Rodpysh (2013)	Third party	Classification	Decision tree, Bayesian network, neural network	Profit function, ROC curve, lift index
	Günther et al. (2014)	Automobile , home, healthcare	Regression	Generalised additive models (GAM)	Receiver operating characteristic (ROC) curve
	Kašćelan, Kašćelan & Novović Burić (2016)	Automobile	Clustering, Regression	K-means clustering, support vector regression (SVR), Kernel logistic regression (KLR)	Absolute error (AE), RMSE, relative error (RE), normalised absolute error (NAE)
	Rose (2016)	Healthcare	Non- parametric machine learning	Artificial neural network, decision tree, random forest	Cross-valided R- square
Fraud detection	Thornton et al. (2014)	Healthcare	Outlier detection	Multivariate analysis, time series, box plot analysis, regression, k- means clustering, trend and peak deviations	Standard deviation
	Kose, Gokturk & Kilic (2015)	Healthcare	Machine learning	Proactive and retrospective analysis, clustering	AUC
	Goleiji & Tarokh (2015)	Automobile	Genetic algorithm method, correlation	Decision tree, Naïve bayesian	Total accuracy percentage


Joudaki et al., (2016)	Healthcare	Classification	Hierarchical clustering	Discriminant analysis
Nian et al. (2016)	Automobile	Spectral ranking analysis	One-class support vector machine (SVM), local outlier factor (LOF), supervised random forest	ROC curve, AUC
Hargreaves & Singhania (2016)	Automobile	Statistical hypothesis testing	Chi-squared test, independent sample t-test	p-value significance

## 4. Conclusions

Data analytics is being implemented worldwide across various industries. Data mining and machine learning techniques have immense potential in offering businesses with a competitive edge over their adversaries. Extensive literature exists on numerous domains where different analytical techniques were applied. The objective of this paper was to conduct a review of the literature on the analytical techniques employed in the insurance industry. Three major areas were discussed namely customer analytics, risk prediction and fraud detection. Finally, a critical evaluation was made based on the literature studied and a summary was provided on the different techniques implemented by the researchers.

### References

- Almana, A., Aksoy, M. & Alzahran, R. (2014) A Survey On Data Mining Techniques In Customer Churn Analysis For Telecom Industry. *International Journal of Engineering Research and Applications*. 4(5). p.165-171.
- Amoo, A., Akinyemi, B., Awoyelu, I. & Adagunodo, E. (2015) Modeling & Simulation of a Predictive Customer Churn Model for Telecommunication Industr. *Journal of Emerging Trends in Computing and Information Sciences*. 6(11). p.630-641.
- Ansari, A. & Riasi, A. (2016) Modelling and evaluating customer loyalty using neural networks: Evidence from startup insurance companies. *Future Business Journal*. 2(1). p.15-30.
- Bănărescu, A. (2015) Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance.* 32(1). p.1827-1836.
- Delafrooz, N. & Farzanfar, E. (2016) Determining the Customer Lifetime Value based on the Benefit Clustering in the Insurance Industry. *Indian Journal of Science and Technology*. 9(1).



- Desik, P., Samarendra, B., Soma, P. & Sundari, N. (2016) Segmentation-Based Predictive Modeling Approach in Insurance Marketing Strategy. *IUP Journal of Business Strategy.* 13(2).
- Fang, K., Jiang, Y. & Song, M. (2016) Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers & Industrial Engineering.* 101. p.554-564.
- Goleiji, L. & Tarokh, M. (2015) Identification of Influential Features and Fraud Detection in the Insurance Industry using the Data Mining Techniques (Case Study: Automobile's Body Insurance). *Majlesi Journal of Multimedia Processing.* 4(3). p.1-5.
- Golmah, V. (2014) A Case Study of Applying SOM in Market Segmentation of Automobile Insurance Customers. *International Journal of Database Theory and Application*. 7(1). p.25-36.
- Goonetilleke, T. & Caldera, H. (2013) Mining Life Insurance Data for Customer Attrition Analysis. *Journal of Industrial and Intelligent Information.* 1(1). p.52-58.
- Günther, C., Tvete, I., Aas, K., Sandnes, G. & Borgan, Ø. (2014) Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*. 2014(1). p.58-71.
- Hanafizadeh, P. & Paydar, N. (2013) A Data Mining Model for Risk Assessment and Customer Segmentation in the Insurance Industry. *International Journal of Strategic Decision Science*. 4(1). p.52-78.
- Hargreaves, C. & Singhania, V. (2016) Analytics for Insurance Fraud Detection: An Empirical Study. American Journal of Mobile Systems, Applications and Services. 1(3). p.227-232.
- Jandaghi, G. & Moradpour, Z. (2015) Segmentation of Life Insurance Customers Based on their Profile Using Fuzzy Clustering. *International Letters of Social and Humanistic Sciences.* 61. p.17-24.
- Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M. & Arab, M. (2016) Improving Fraud and Abuse Detection in General Physician Claims: A Data Mining Study. *International Journal of Health Policy and Management.* 5(3). p.165-172.
- Kašćelan, V., Kašćelan, L. & Novović Burić, M. (2016) A nonparametric data mining approach for risk prediction in car insurance: a case study from the Montenegrin market. *Economic Research-Ekonomska Istraživanja*. 29(1). p.545-558.
- Kose, I., Gokturk, M. & Kilic, K. (2015) An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Applied Soft Computing.* 36. p.283-299.
- Li, Y., Yan, C., Liu, W. & Li, M. (2016) *Research and Application of Random Forest Model in Mining Automobile Insurance Fraud.* In Proceedings of the 12th International



Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. IEEE.

- Mohammadi, V., Albadvi, A. & Teymorpur, B. (2014) Predicting Customer Churn Using CLV in Insurance Industry. *Shiraz Journal of System Management.* 2(1). p.39-49.
- Muhammad, S. (2014) Fraud: The Affinity of Classification Techniques to Insurance Fraud Detection. *International Journal of Innovative Technology and Exploring Engineering*. 3(11). p.62-66.
- Nian, K., Zhang, H., Tayal, A., Coleman, T. & Li, Y. (2016) Auto Insurance Fraud Detection using Unsupervised Spectral Ranking for Anomaly. *The Journal of Finance and Data Science.* 2(1). p.58-75.
- Polyakova, M. (2016) Risk Selection and Heterogeneous Preferences in Health Insurance Markets with a Public Option. *Journal of Health Economics*. 49. p.153-168.
- Rahman, M., Arefin, K., Masud, S. & Sultana, S. (eds.) (2017) Analyzing Life Insurance Data with Different Classification Techniques for Customers' Behavior Analysis. In Advanced Topics in Intelligent Information and Database Systems. 1st ed. Springer International Publishing, p.15-25.
- Ravasan, A. & Mansouri, T. (2015) A Fuzzy ANP Based Weighted RFM Model for Customer Segmentation in Auto Insurance Sector. *International Journal of Information Systems in the Service Sector*. 7(2). p.71-86.
- Rawte, V. & Anuradha, G. (2015) *Fraud Detection in Health Insurance using Data Mining Techniques.* In International Conference on Communication, Information & Computing Technology. IEEE.
- Robson, J. (2015) General Insurance Marketing: A Review and Future Research Agenda. *Journal of Financial Services Marketing.* 20(4). p.282-291.
- Rodpysh, K. (2013) Providing a Method for Determining the Index of Customer Churn in Industry. *International Journal of Information Technology, Control and Automation*. 3(1). p.61-70.
- Rodriguez, S. & Shin, H. (2013) Developing Customer Churn Models for Customer Relationship Management. *International Journal of Business Continuity and Risk Management.* 4(4). p.302.
- Roodpishi, M. & Nashtaei, R. (2015) Market Basket Analysis in Insurance Industry. *Management Science Letters*. 5(4). p.393-400.
- Rose, S. (2016) A Machine Learning Framework for Plan Payment Risk Adjustment. *Health Services Research*. 51(6). p.2358-2374.
- Sithic, H. & Balasubramanian, T. (2013) Survey of Insurance Fraud Detection Using Data Mining Technique. *International Journal of Innovative Technology and Exploring Engineering.* 2(3). p.62-65.



- Sivarajah, U., Kamal, M., Irani, Z. & Weerakkody, V. (2017) Critical Analysis of Big Data Challenges and Analytical Methods. *Journal of Business Research.* 70(C). p.263-286.
- Thornton, D., van Capelleveen, G., Poel, M., van Hillegersberg, J. & Mueller, R. (2014) *Outlier-based Health Insurance Fraud Detection for U.S. Medicaid Data.* In Proceedings of the 16th International Conference on Enterprise Information Systems. Scitepress. p.684-694.
- Umamaheswari, K. & Janakiraman, D. (2014) Role of Data mining in Insurance Industry. *An International Journal of Advanced Computer Technology.* 3(6). p.961-966.
- Vafeiadis, T., Diamantaras, K., Sarigiannidis, G. & Chatzisavvas, K. (2015) A Comparison of Machine Learning Techniques for Customer Churn Prediction. *Simulation Modelling Practice and Theory.* 55. p.1-9.
- Wang, S., Pai, H., Wu, M., Wu, F. & Li, C. (2017) The Evaluation of Trustworthiness to Identify Health Insurance Fraud in Dentistry. *Artificial Intelligence in Medicine.* 75. p.40-50.
- Zhou, L. & Chen, Q. (2014) Customer Segmentation, Return and Risk Management: An Empirical Analysis Based on BP Neural Network. *Journal of Chemical and Pharmaceutical Research*. 6(6). p.698-703.



# **Wireless Power Supply for Portable Devices**

Ahmed Najib Bhutta School of Engineering Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia Email: deathshead747@gmail.com

Veeraiyah Thangasamy School of Engineering Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia Email: dr.veeraiyah@apu.edu.my

Chandrasekharan Nataraj School of Engineering Asia Pacific University of Technology & Innovation 57000 Kuala Lumpur, Malaysia Email: chander@apu.edu.my

*Abstract*— As consumer electronics, smartphones, biomedical implants, electric vehicles etc. become more pervasive, there needs to be a paradigm shift for overcoming regular charging and to lessen the dependency on batteries. Wireless Power Transfer (WPT) is poised to surmount traditional charging methods and usher in a new era of IoT (Internet of Things) devices. However, the current state of WPT is riddled with proximity issues, poor efficiency, mediocre transmission distances, excessive heat-up and design complexions. This research aims to enhance the efficiency and the effective transmission distance. A WPT system has been designed based on a mixed, LCC topology along with a highly-efficient Class-E Power Amplifier as the source. The overall transfer-link efficiency attained for the designed WPT system is 90.026%. A multi-dimensional coil hierarchy can further be adopted for the designed WPT system for achieving spatial freedom. Furthermore, several tests have been conducted to test the system's overall feasibility and to enhance its performance even more.

*Index Terms*—wireless power transmission, class-E Power Amplifier, GaN FET, 13.56 MHz, Magnetic Resonance Coupling, Oscillator, Intermediate Coils.



### 1. Introduction

O ne of the major driving forces that has brought about accelerated development of society, as well as enhanced the overall state of life, is Energy. The sudden increase in the advancement of science and technology, has enabled the world to become more affluent with a vast variety of portable consumer electronics, biomedical, vehicular and industrial devices. This unanticipated surge of portable devices has given rise to a new array of challenges; The foremost issue being, the high level of dependency on batteries for operational purposes. Battery technology has pretty much failed to keep up with the ever-dynamic consumer electronic industry, with the advances becoming almost stagnant.

With the advent of power electronics, wireless power transfer (WPT) or wireless charging technology has quickly progressed and is gaining popularity due to its simplicity & efficiency. Most current generation WPT systems, rely on Inductive Coupling as a means for efficiently charging devices. But vastly degrading efficiency, as a result of misalignment and angular placement, defeats the entire essence of going wireless [1]. Moreover, these inductive based WPT systems tend to suffer from intrinsic winding losses that leads to slow charging times and excessive heat dissipation.

The other type of WPT technique for efficiently transmitting over large distances is based on RF-Radiation, which as the name suggests is not suitable everywhere and poses a safety hazard, for users with biomedical implants. However, research conducted by MIT has painted magnetic resonant coupling to be quite an effective solution, due to its high-power transfer efficiency, covering a span of several meters (medium range). This new entrant has the potential for possibly commercializing midrange wireless power transfer systems [2].

Although magnetic resonance has a substantial edge when it comes to transmission distance, in contrast to electromagnetic induction, this technology has inherent limitations, which are most visible when there is even a slight variation in the operating parameters. Furthermore, a small difference in these parameters & resonance frequencies, will significantly impact the transmission performance and lead to frequency splitting for changing coupling co-efficient [3].

Whilst creating such WPT systems, the source used for powering the entire system often goes overlooked. However, according to a number of researches, the overall efficiency of the system can be enhanced by improving the transmission side of the WPT system [4]. Thus, a Class-E Power Amplifier (PA) has also been designed to be incorporated into the system, to improve its overall efficiency. A Class-E PA has been specifically selected due to its extreme switching and operational capabilities, allowing the PA to reach a theoretical operational efficiency of 100%. For this research, a GaN FET has been utilized because of its high switching rates and to reduce the overall size of the proposed WPT system [5].



## 2. WPT System Design & Construction

#### 2.1 Block Diagram

As can be observed from Fig. 1, the block diagram of the entire system is separated into two main parts, the transmission side (left) and the receiving side (right). At the transmission side, it is apparent that the Class-E Power Amplifier acts as the source, which powers up the entire system and eventually leads to charging the device attached. After the amplification, the power coil transfers the energy to the transmitter coil. The transmitter coil, which is tuned to operate at the resonance frequency, magnetically transfers power to the receiving coil.



Fig. 1: Operational Block Diagram of the WPT System

At the receiving end, the power transmitted is accepted by the receiving coil through the phenomenon of Magnetic Resonance Coupling; It too is set to operate at the resonant frequency. The receiving coil then further transfers the power to the final load coil. The load coil can either be directly hooked to a device (load) or it can be attached to a rectifier first. The rectifier smoothens out the received voltage, which is then supplied to the attached device for a more consistent charging experience.

#### 2.2 Class-E Power Amplifier

The Class-E PA has been selected to power the proposed WPT system. A typical Class-E Amplifier configuration includes a transistor that acts as a switch, which is further attached to an RF choke (L1) at the drain, an LC resonator (L2 & C2) connected in series and a resistive load (R1) as shown in Fig. 2. As a means for making the circuit more compact, many researches tend to exclude the shunt capacitor (C1), attached in parallel, but this significantly impacts the efficiency [5]. Since most portable devices such as



smartphones, usually require only about 12W and 2.4A for fast-charging purposes, the supply & drain voltages have been set to 5V.

For this research, a high efficiency Gallium Nitride (GaN FET) based switchingmode transistor has been opted for, instead of the more traditionally used silicone MOSFET. By incorporating a Gallium Nitride (GaN) enhancement mode FET in the Class-E PA, the power density provided is significantly enriched, as it tends to operate at higher temperatures and considerably high switching frequencies. Furthermore, it is also capable for providing amplification, for a multitude of varying loads at the highest efficiency [4].



Fig. 2: Designed Class-E PA with GaN FET attached.

All of the component values are derived by making use of certain formulae, which are provided below [5]:

$$R1 = \frac{0.577 V dc^2}{P_{out}}$$
(1)

Where the load resistance can be calculated by setting an output power (12W) and supplying an input gate voltage of 5V. Using the R1 value obtained, the rest of the components are calculated accordingly:

$$C1 = \frac{1}{5.447\,\omega R} \tag{2}$$

$$L1 = \frac{1.525R}{\omega}$$
(3)

However, the LC Resonator values are calculated by including the proposed WPT's Q-factor, using the following formulae [5]:



$$C2 = \frac{1}{\omega QR}$$

$$L2 = \frac{1}{\omega^2 C2}$$
(4)
(5)

The values for each component was calculated using these formulas and all of the design parameters for the Class-E PA are provided in Table 1.

Components	Values
Frequency	13.56 MHz
V1	5V
V2	5V
L1	16.23nH ≈ 16nH
C1	1795.65pf ≈ 1800pf
L2	17.466nf ≈ 17.5nF
C2	7.887nH ≈ 8nH.
R1	1.20Ω

**Table 1**: Design Parameters for the Class-E Power Amplifier.

#### 2.3 4-Coil WPT System

The proposed Wireless Power Transfer (WPT) System has been designed to further enhance the efficiency and the distance of transmission. Observing Fig. 3, the system designed is unlike other traditional WPT systems, in the way that it utilizes the 4-Coil regime for it to function. This system performs on the basic principles of magnetic resonance, with the added benefits of the intermediate, Tx & Rx coils [3]. Furthermore, the design topology has been tweaked from the conventional series-series to the mixed LCC topology.



Fig. 3: Proposed 4-Coil SCMR WPT System.



The transmission stage is comprised of both the power coil and the transmitter coil. After the power is amplified, it is ready to be transferred that is carried out by the power coil, which is essentially two capacitors connected in parallel to an inductor. The capacitors attached, work in conjunction with the inductor, as a means for providing a simpler solution to impedance matching. The power coil is linked to the transmitter coil magnetically that is represented on the schematic (Fig. 3), as the coupling co-efficient (k), which has been set to 0.15, for the linkage between coils 1 & 2 (power & transmitter).

Here the value of 'k' not only represents, how tightly the coils are linked to one another but also acts as a function of distance between the power coil and the transmitter coil. Generally, the coupling co-efficient value diminishes when the distance is increased [4]. Due to the magnetic link present, the power is transferred to the transmitter coil. The Tx coil also functions on the same principles as the power coil and has been attached in a similar fashion. Being set at the resonant frequency, it is able to efficiently transmit the evanescent power-carrying waves to the receiving side.

The values for all the RLC components have been premeditated to operate at the selected resonance frequency of 13.56 MHz. The following calculations have been performed, based on the design parameters (Table 2), and are calculated using the following formulas.

Wheeler's Formula for calculating the value of the coil:

$$L = \frac{N^2 + R^2}{2.54(9R + 10H)}$$
(6)
where, L = inductance (uH)

N = number of turns of wire

R = radius of coil (cm) H = height of coil (cm)

Since the resonant frequency has already been set at the source, the following formula may be used to determine the capacitance value:

$$f_r = \frac{1}{2\pi\sqrt{L(C1+C2)}}$$
(7)

where, Fr = resonance frequency

L = inductance (uH) C1 = series capacitance (pF) C2 = parallel capacitance (pF)

Furthermore, the receiving end is comprised of the receiver & load coils, respectively and are arranged in the aforementioned manner. However, the link (k) between the transmitter and receiving coil has been set to 0.043, as can be seen from Fig. 3. This has been done for further increasing the actual distance between the power and load coils, as well as to overcome frequency splitting and improve the overall efficiency of the entire system [4].



Coil	Ν	R	Н	L	F	C1	C2
(inductor)	(turns)	(cm)	(cm)	(uH)	(MHz)	(pF)	(pF)
Power	2	5	3.3	0.5	13.56	150	125.52
Transmitter	3	6	4.4	1.3	13.56	75	30.97
Receiver	1.67	6	4.4	0.4	13.56	125	219.4
Load	1	3.7	2	0.1	13.56	277	1100

Table 2: Design Parameters for the WPT System.

### 3. Working Principle

The operational working principle of the proposed WPT system is shown in Fig. 4. When a DC voltage is supplied to the Class-E Amplifier, it goes through the choke inductor (L1) and supplies a constant DC current to the GaN FET (EPC 2022) used. This transistor operates as switching mode amplifier and allows an AC current to flow through the transistor, when it is switched on. When the transistor switches OFF, the current is then rerouted to the load, producing a voltage. In an ideal scenario, this transition is seamless with a theoretical efficiency of 100%. However, power dissipation at harmonic frequencies, negates the efficiency, which is why an LC resonator is incorporated to regulate such harmonics. In the end, the amplified power is provided to the load and in this case, to the power coil for further transmission [5].

After being injected with the amplified power, the power coil begins to oscillate at the resonant frequency, to transfer the power to the intermediary Tx coil. When certain frequencies are applied, many objects tend to vibrate, producing greater amplitudes than at other frequencies, this phenomenon is known as resonance [6]. Thus, for this entire system the resonant frequency is set at 13.56 MHz.

The transferred power from the power coil is fed in to the intermediate Tx coil, which itself is already vibrating at the resonant frequency, due to the aid of impedance matching capacitors. Impedance matching is a necessary framework for efficient Wireless Power Transfer, as it optimally tunes the coils on both the transmitting and receiving ends, to resonate at a particular frequency, for maximum exchange and minimal losses [2].

The Tx coil then further transmits the power to the Rx coil magnetically. Both of these coils are the intermediate coils that have been inserted to make the WPT system more efficient. These coils are free from source internal resistances & other parasitic losses, leading to an improvement in the overall Q-factor [3]. The receiver coil again oscillates to finally transmit the power to the load coil. After having received the power transmitted all the way from the amplification and power coil stage, the load coil is connected to the portable device (load). It can alternatively be connected to a rectifier first, that will ensure a steady amount of current is provided to replenish the device.





Fig. 4: Operational Principle Flowchart of designed WPT system.



## 4. Simulation Results

### 4.1 Class-E Power Amplifier

The Class-E Power Amplifier has been designed using LTspice due to a lack of a GaN FET in ADS. The PA has been setup as shown in Fig. 2 and its transient behaviour is analysed. As can be observed from Fig. 5, the voltage at the output is around 7.3V, with the current across the load being 6.2A. A drain input voltage of 5V has been provided and the drain current that exists is 9.4A, as can be evident from Fig. 6. Hence, the total efficiency of the designed Class-E Power Amplifier obtained is 95.75%, which is a bit lower than the theoretical efficiency of 100%.



Fig. 5: Output voltage & current waveforms for the Class-E PA



Fig. 6: Drain current & gate voltage with respect to time for the Class-E PA.



#### 4.2 Wireless Power Transfer System

The WPT system has been setup with the S-Parameters terminations for testing the transfer-link efficiency of the system and the results are shown in Fig. 7 through Fig. 9. As can be observed from Fig. 7, the S11 magnitude obtained for the designed WPT system is just 0.252 and should be relatively low for an efficient system. The S21 parameter obtained is quite important as it signifies the ratio of the receiver's receiving power divided by the transmitter's transmitting power and hence, the total power that can be transferred [3]. Fig. 8, illustrates the value of S21, which is 0.949, with 1 being the maximum amount of power transferred. Finally, as can be observed from Fig. 9, by performing the S-parameters simulation, the overall transfer-link efficiency of the designed WPT system is 90.026%.



Fig. 7: S11 Magnitude result of the designed WPT system.



Fig. 8: S21 Magnitude result of the designed WPT system.



Fig. 9: Transfer-link efficiency of the designed WPT system.



## 5. Testing & Evaluation

In order to assess the overall performance and efficiency of the designed WPT systems, a number of tests were carried out, which are as follows:

### 5.1 Type of Topology

One of the most vital design changes that adversely affects the performance and link transfer efficiency of WPT systems, is the type of topology used. Conventionally, most inductive-coupling based charging systems make use of the series-series or LC compensation topology [4]. However, for the final design a mixed LCC type compensation topology was used and its performance was evaluated with the series-series one.

erformance comparison data between unterent types of t					
Topology	S11	S21	Efficiency		
Series-Series	0.973	0.184	3.371%		
Mixed LCC	0.252	0.949	90.026%		

Table 3: Performance comparison data between different types of topologies.

As can clearly be observed from Table 3, using the traditional series-series topology without changing the design parameters, greatly degrades the entire performance of the system, when compared to the mixed LCC topology.

### 5.2 Effects of the Coupling Co-efficient

Fundamentally, the existing amount of coupling for an inductive system can be expressed as a fraction number between 0 and 1. Where 0.5<k<1 represents a tightly coupled system and 0<k<0.5 signifies a loosely coupled system. This is one of the most important factors that effects the transfer-link efficiency and determines the overall performance of the WPT system [2]. The coupling co-efficient is effectively determined by two factors, which is the distance between coils and their relative sizes [3].

e						
	Coupling co-	S21 Magnitude	Transfer-Link			
	efficient Value		Efficiency			
	K23 = 0.03	0.864	76.229%			
	K12 = 0.1	0.861	73.464%			
	K34 = 0.1	0.891	79.959%			

Table 4: Performance evaluation for varying Coupling Co-efficient (K)

As can be seen from Table 4 above, reducing the distance between the transmitter and receiver (k23 = 0.03), whilst keeping k12 & k34 fixed, the efficiency of the system decreases significantly and the frequency spectrum becomes narrower. Furthermore,



reducing the value of k12 to 0.1, the efficiency of the system is drastically impacted and frequency splitting occurs. Frequency splitting is a phenomenon at which the efficiency of a system, peaks at both below & above the resonant frequency [4]. Finally, when coupling between the receiver coil & load coil (k34) is increased to 0.1, while keeping k12 & k23 fixed, the efficiency decreases and the effects of frequency splitting are reduced.

### 5.3 Effects of using different System Types

A 4-Coil WPT system was designed so as to enhance the efficiency and increase the coverage area (distance) of the system [3]. The same system was recreated as a 2-Coil & 3-Coil system and its effects were analysed in detail as shown in Table 5.

System Type	k	S21	Efficiency		
Two-Coil	0.5	0.707	49.974%		
Three-Coil	0.5	0.810	65.589%		
Designed 4-Coil	0.043	0.949	90.026%		

 Table 5: Efficiency comparison between 2-Coil, 3-Coil and the designed WPT system.

It can clearly be seen that even after substantially reducing the distance between the coils for the 2 & 3-Coil WPT setup, the power transfer ratio and the transfer-link efficiency takes a hit, leading to poor efficiency.

### 6. Conclusion

A wireless power supply for powering portable devices was designed and its performance was tested. A 4-Coil design WPT, with mixed LCC topology and highly efficient Class-E power amplifier was designed in this research. The 4-Coil design was incorporated to benefit from the added intermediate coils, which not only aid in increasing the transmission distance but also to mitigate source internal resistances. This type was later evaluated against existing solutions and proven to be the most suitable. A well-balanced link was achieved between the coupling co-efficient that increased the distance without hampering the efficiency. Furthermore, the designed Class-E PA was able to achieve an efficiency of 95.75%, a little short of the theoretical 100% efficiency. By using the mixed LCC topology, along with all of the design factors, the WPT system with a power transfer of 0.949 and an overall transfer-link efficiency of 90.026% was achieved. The addition of intermediate coils, allows the design to be more modular and the size of the receiver coil can further be reduced, for powering biomedical implants.



### References

- [1] S. Ho, J.Wang, W. Fu and M.Sun, "A Comparative Study Between Novel Witricity and Traditional Inductive Magnetic Coupling in Wireless Charging", *IEEE Transactions on Magnetics*, 47 (5), p. 1522-1525, 2011.
- [2] X. Lu, P. Wang, D. Niyato, D. Kim and Z. Han, "Wireless Charging Technologies: Fundamentals, Standards, and Network Applications", *IEEE Communications Surveys AND Tutorials*, 18 (2), 2016, pp. 1413-1452.
- [3] A. Sah, "Design of Wireless Power Transfer System via Magnetic Resonant Coupling at 13.56MHz" *Proceedings of 10E Graduate Conference, Kathmandu: Institute of Engineering*, 2013, pp. 372-385.
- [4] A. Sah and D. Pant, "Analysis and Optimization of Wireless Power Transfer Link", *IOE Graduate Conference, Kathmandu: Institute of Engineering*, 2014, pp 269-278.
- [5] M. Ali and H. Nugroho, "Effective power amplifier of wireless power transfer system for consumer electronics," *IEEE International Conference on Power System Technology (POWERCON)*, Wollongong, NSW, 2016, pp. 1-5.
- [6] S. D. Barman, A. W. Reza, N. Kumar, M. E. Karim and A. B. Munir, "Wireless powering by magnetic resonant coupling: Recent trends in wireless power transfer system and its applications" *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1525–1552, 2015.





