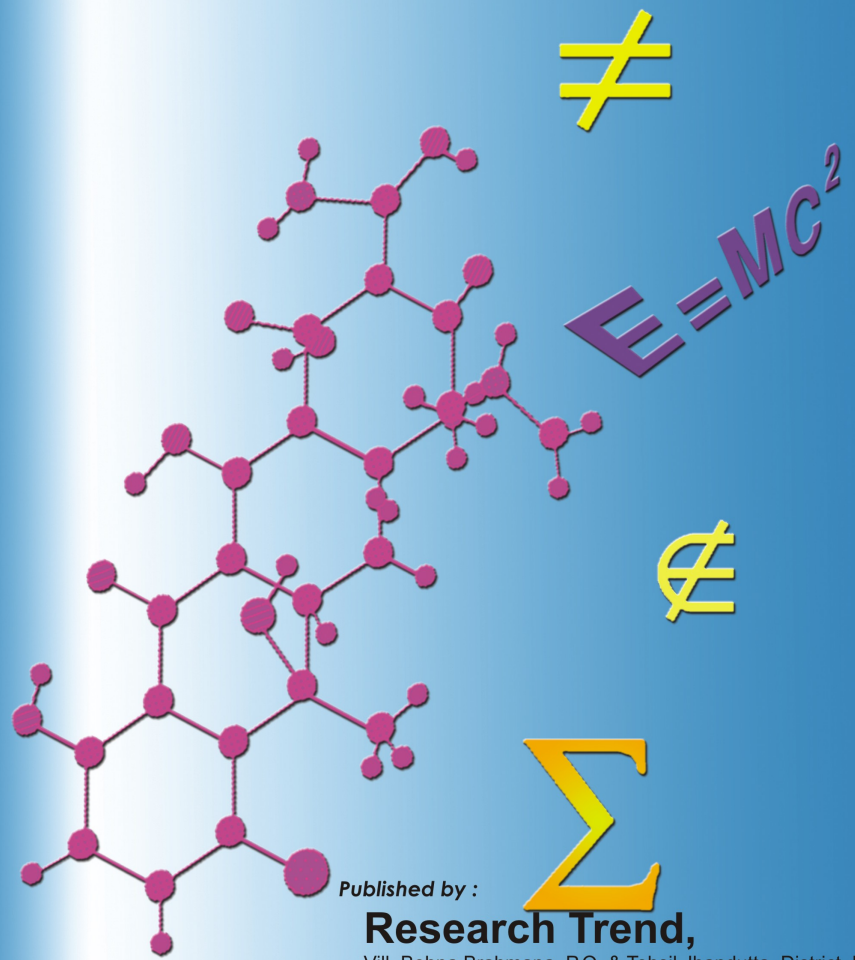
International Journal of Theoretical & Applied Sciences

VOL. 10(1a): Special Issue, January-June, 2018

Print ISSN No.: 0975-1718

Online ISSN No.: 2249-3247



Published by :

Research Trend,

Vill. Behna Brahmana, P.O. & Tehsil Jhandutta, District. Bilaspur, Himachal Pradesh-174031 (India) Website: www.researchtrend.net

Special Issue Released

for

International Conference on Emerging Social, Mobile, Analytics and Cloud

e-SMAC 2018

24th January 2018

SMAC

Social Mobile Analytics Cloud

Organised by

Department of Computer Science

SRI RAMAKRISHNA

COLLEGE OF ARTS AND SCIENCE FOR WOMEN

Affiliated to Bharathiar University | Approved by AICTE

Accredited by NAAC and An ISO Certified Institution

Coimbatore - 44

Vol. 10(1a): 2018 Special Issue Jan.–June, 2018 ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

***Research Trend*,**
J-4/50, 2nd Floor, Khirki Extension, Malviya Nagar, New Delhi-17.
Website: www.researchtrend.net **Email**: dheeraj_vasu_72066@yahoo.co.in,
researchtrend09@gmail.com, Mobile : **9868001440**

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): (2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Editorial

"India, a land which gave birth to civilization in ancient times and where much of the earlier tradition and wisdom guides actions even in modern times the philosophy of Vasudhaiva Kutumbakam which means that the whole universe is one family, dominates global efforts to protect the global commons".

We are profoundly privileged to bring before you the efforts of a few genius minds in the form of ***"International Journal of Theoretical and Applied Sciences".*** Its aims to encourage to make aware the human being towards scientific attitude for the betterment of ecosystem and social life.

Cheers to all those involved directly on front or indirectly behind the curtain in this noble attempt of serving ecosystem.

Science belongs to the whole world, and before it, vanish all the barriers of nationality. With the resonance of science in all the activities of our lives, we are trying to marvel this age of specialization. Standing on this verge of eternity we are trying to possess more and more of power and pelf. For it, we require a quality of mind, which should be special and should have an extreme advantage in leading to make discoveries. What matters is the power of never letting exceptions go unnoticed.

The foundation blocks of research are to do the right thing, at the right time, in the right way; to anticipate requirements; to develop resources and then to recognize no impediments and thence to master circumstances. One has to act from reason rather than rule and to be satisfied with nothing short of perfection. True researcher resides in the capacity or evaluation of uncertain, hazardous and conflicting information. Curiosity, Confidence, Courage and Constancy are the hallmarks. Research is a long road to be traded by the brave ones. One has to brave all odds, long pangs of suffering and frustration to bring the work in hand to fruition. To attain the pinnacle of success one ought to nurture research with hard work and toil.

We congratulate and wish luck to the researchers for their contributions and aspire that these works will leave a glowing trail for the generations to come. These works will act as lighthouses to the future generations and rare milestones in the fields of Science and Technology. It may also provide the courage to tread the long and weary path of research, as success and hard work has a taste beyond everything.

– Editor-in-chief

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): (2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Table of Contents Special Issue Vol. 10(1a): 2018

International Journal of Theoretical & Applied Sciences, Special Issue 10(1a): 01-07(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Research and Developments on Measurements of Hairiness in Cotton Yarn by Image Processing based Techniques

R. Sudha[1] and Dr. V. Chitraa[2]
[1]Ph.D. Scholar, Department of Computer Science,
CMS College of Science and Commerce, Coimbatore (Tamilnadu), India
[2]Associate Professor, Department of Computer Science,
CMS College of Science and Commerce, Coimbatore (Tamilnadu), India

ABSTRACT: The hair projecting out from the main body of the yarn are called Yarn Hairiness. It is one of the main aspects, key indicator of limitation to detect the yarn quality. It affects the appearance of yarn and succeeding handing out of textile process. It is in most state of affairs an undesirable property, giving rise to difficult of fabric production and also worsens the fabric appearance. Many growths regarding various researches on yarn hairiness have been described. These researchers handle the various aspects such as measurement of hairiness, exhibiting, imitation, spinning adaptations, post spinning managements and new advanced technological solution idea to automatically quantify yarn to reduce hairiness and many techniques for detecting .by using image processing the yarn associated fabrics prediction are carried out, and by means of Artificial Intelligence techniques in detection too. This review is an attempt to systematically give all significant recent growth and advancements regarding yarn hairiness, further possibilities of research and future work are also briefly deliberated.

Keywords: Artificial Intelligence, Hairiness Measurement, Hairiness Modeling, Yarn Spinning, Image Processing, Segmentation, Yarn hairiness.

I. INTRODUCTION

Yarn hairiness is fiber head or fiber tail that exposed to yarn stem, which is caused by being not twisted completely [2-5]. The basic form of hairiness is divided into two categories: one is the protruding fiber ends; another one is the looped fibers arched out the yarn stem. The whole body is attached to the surface of the yarn stem and not involved in the yarn, called 'floating hairiness', which is not regarded as yarn hairiness. The schematic diagram is shown in Fig. 1. [1]

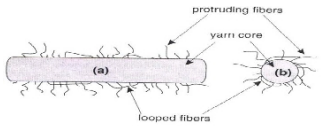


Fig. 1.

Yarn hairiness is a critical yarn quality parameter. There has been significant interest in minimizing yarn hairiness to improve both yarn and fabric quality. The long yarn hairiness (≥ 3 mm) in yarn hairiness is more, the appearance of yarn is coarser, and the hairiness is more entangled in the process of weaving, so it is important to accurately measure the long yarn hairiness [1]. Various methods of hairiness testing have been developed and evaluated. Attempts to model and predict yarn hairiness with different techniques have also been made over the last few years. This study is an effort to continue the series of reviews on yarn hairiness previously conducted by Noman Haleem and Xungai Wang. The last review on this topic was published by Noman Haleem and Xungai Wang in 2014 with a focus on various Improvements in yarn hairiness measurement, development of new spinning systems, hairiness reducing mechanisms, and post-spinning processes affected by yarn hairiness [6,7]. Since then, tremendous progress has been made in the field of yarn hairiness, ranging from modified spinning systems to applications of advanced modeling and measurement techniques to precisely determine true yarn hairiness. Over 600 research studies directly or indirectly related to yarn hairiness have been reported in the last decade from which at least 300 are directly focused on yarn hairiness. Even long-established theories on the fundamental characteristics of yarn hairiness are now challenged by new experimental evidence. It is now opportune to provide another update on yarn hairiness. This paper brief the importance and cause of yarn hairiness in the second section and third sections.it also discuss on the measurement of yarn using sensor and signal processing in the fourth sections and in the fifth section the application of image processing in the measurements in elaborately discussed in the following sections sixth section and limitation and conclusions in the preceding sections.

II. IMPORTANCE OF YARN HAIRINESS

It is in most circumstances an undesirable property, giving rise to problem of fabric production and also deteriorates the fabric appearance. Measurement of hairiness dependent on the method chosen for detecting the hairs. Fibers protruding out from the main body of the yarn are called hairiness. In most circumstances it is an undesirable property, giving rise to problems in fabric production. In natural short staple fiber yarns, the reason behind being wide differences in fiber thickness, maturity and inadequate spinning process. In filament yarns, hairiness occurs due to weak monofilaments, inadequate finish, rough surfaces, loose running monofilaments which break at subsequent process (due to loose run, denier is reduced hence weaken yarn). It is not possible to represent hairiness with a single parameter because the number of hairs and the length of hairs both vary independently. A yarn may have a small number of long hairs or a large number of short hairs or any combination in between. The problem is then which combination should be given a higher hairiness rating. It is considered that there are two different exponential mechanisms in operation, one for hairs above 3mm long and one for these below. The number of hairs exceeding 3mm in length as a percentage of the total number of hairs is found to be linearly related to the linear density of the yarn.

III. CAUSES AND FACTS BEHIND YARN HAIRINESS

Yarn hairiness is a necessary consequence of twisting Finite length fibers into a yarn in spinning process. However, the level of hairiness is affected by the properties of the raw material, yarn parameters, and processing parameters and various changes in the Spinning Process. This may leads to directly or indirectly cause hairiness in yarn.

Pillay proved that there is a high correlation between the number of protruding ends and the number of fibers in the yarn cross-section. Torsion rigidity of the fibers is the most important single property affecting yarn hairiness. Other factors are flexural rigidity, fiber length and fiber fineness. Mixing different length cottons-No substantial gain in airiness. Although the hairiness of a yarn could be reduced to some extent by the addition of a longer and finer cotton to the blend. The extent of reduction is not proportional to the percentage of the longer and finer component. This is probably due to the preferential migration of the coarser and shorter component, which has longer protruding ends, from the yarn body. The addition of wastes to the mixing increases the yarn hairiness; the effect of adding comber waste is greater than that of adding soft waste. Blending-not a solution to hairiness.

The blended yarns are rather more hairy than expected from the hairiness of the components; a result similar to that found in cotton blends.

This may be due to the preferential migration of the shorter cotton fibers; a count of the number of protruding ends of both types of fiber shows that there is more cotton fiber ends than expected, although the difference is not very great. The number of protruding ends is independent of twist, whereas the number of loops decreases when the yarn twist increases because of a greater degree of binding between the fibers owing to twist. The number of wild fibers decreases only very slightly with twist because of their position on the yarn periphery. The proportion of fiber ends that protrude from the yarn surface, counted microscopically has been found to be about 31% of the actual number of ends present in the yarn. If the length of the protruding fiber ends as well as that of the loops is considered, the mean value of the hairiness increases as the cross-sectional area increases and decreases with the length of the loops. The hairiness is affected by the yarn twist, since an increase in twist tends to shorten the fiber ends.

Wild fibers are those for which the head alone is taken by the twist while the tail is still gripped by the front drafting rollers. Fiber length influences hairiness in the sense that a greater length corresponds to less hairiness. Cotton yarns are known to be less hairy than yarns spun from man-made fibers. The possible reason for this is the profile of the two fibers. Because of taper, only one end, the heavier root part of the cotton fiber, ends to come out as a protruding end in a cotton yarn. With man-made fibers, both ends have an equal probability of showing up as protruding ends. If the width of the fiber web in the drafting field is large, the contact and friction with the bottom roller reduce the ability of the fibers to concentrate themselves and hairiness occurs. This effect is found more in coarse counts with low TPI. This suggests that the collectors in the drafting field will reduce yarn hairiness.

The yarn hairiness definitely depends on the fibers on the outer layer of the yarn that do not directly adhere to the core. Some of them have an end in the core of the yarn gripped by other fibers whereas others, because of the mechanical properties of the fiber (rigidity, shape, etc.) emerge to the surface. During the twisting of the yarn, other fibers are further displaced from their central position to the yarn surface. Greater the fiber parallelization by the draw frame, lower the yarn hairiness. An increase in roving twist results in lower yarn hairiness, because of smaller width of fiber web in the drafting field. The number of fiber ends on the yarn surface remains fairly constant; the number of looped fibers reduces in number and length on increasing twist.

Combed yarn will have low yarn hairiness, because of the extraction of shorter fibers by the comber. Yarn hairiness increases when the roving linear density increases. Yarn spun from double roving will have more hairiness than the yarn spun from single roving. This is due to the increased number of fibers in the web and due to higher draft required to spin the same count.

Drafting waves increase hairiness. Irregularity arising from drafting waves increases with increasing draft. Yarn hairiness also may be accepted to increase with yarn irregularity, because fibers protruding from the yarn surface are more numerous at the thickest and least twisted parts of the yarn. Higher spindle speed – high hairiness. When yarns are spun at different spindle speed, the centrifugal force acting on fibers in the spinning zone will increase in proportion to the square of the spindle speed, causing the fibers ends as they are emerging from the front rollers to be deflected from the yarn surface to a greater extent. Further, at high spindle speed, the shearing action of the traveler on the yarn is likely to become great enough to partially detach or raise the fibers from the body of the yarn. As against the above factors, at higher spindle speeds the tension in the yarn will increase in proportion to the square of the spindle speed, and consequently more twist will run back to the roller nip, so that it is natural to expect that better binding of the fibers will be achieved. The increase in hairiness noticed in the results suggests that the forces involved in raising fibers from the yarn surface are greater than those tending to incorporate them within the body of the yarn at higher spindle speeds.

Higher draft before ring frame-less hairiness. There is a gradual reduction of hairiness with increase in draft. In other word, as the fiber parallelization increases hairiness decreases. Reversing the card sliver before the first drawing head causes a reduction in hairiness, the effect being similar to that resulting from the inclusion of an extra passage of drawing. Smaller roving package-less hairiness. Yarn hairiness decreases with decrease in roving (doff) size, and yarn spun from front row of roving bobbins is more hairy and variable as compare to that spun from back row of rowing bobbins. It may be noted that though the trends are consistent yet the differences are non significant. The spinning tension has a considerable influence on the yarn hairiness. The smaller the tension, the greater the hairiness. This is the reason why heavier travelers result in low yarn hairiness If the traveler is too heavy also, yarn hairiness will increase.

Yarn hairiness is caused by protruding ends, by the presence of a majority of fiber tails. This suggests that these tails will become heads on unwinding and that friction to which the yarn is subjected will tend to increase their length. It is therefore logical that a yarn should be more hairy after winding.

Repeated windings in the cone winding machine will increase the yarn hairiness and after three or four rewinding's, the yarn hairiness remain same for cotton yarns. Winding speed influences yarn hairiness, but the most important increase in hairiness is produced by the act of winding itself. Because of winding, the number of short hairs increases more rapidly that the number of long hairs. In two-for-one twisters (TFO), more hairiness is produced because, twist is imparted in two steps Yarn hairiness also depends upon the TFO speed, because it principally affects the shortest fiber ends. Hairiness variations in the weft yarn will result in weft bars.

IV. MEASUREMENT OF YARN HAIRINESS

There are many hairiness measurement techniques are followed, in that two main instruments based on different working principles are in commercial use. One method works through a sensor array that measures the length of protruding fibers from the yarn core (up to distance) and categorizes them by length. The Second method works through a light source incident on the yarn core, and the amount of light scattered by the protruding fibers is used to work out a hairiness index value for the yarn. Novel solutions proposed for yarn hairiness measurement are largely based on digital image processing because of its ability to provide an intelligent vision system closely resembled to the human eye. Yarn hairiness measurement by means of signal processing is also reported by several researchers, which is discussed in subsequent sections [8].

V. APPLICATION OF SIGNAL AND SENSOR PROCESSING

There are various electronic sensor based systems are used for yarn hairiness measurement. The hardware consisted of lenses, photodiodes, optical filter.[9] Publications per year that have directly or indirectly addressed yarn hairiness diaphragm. The yarn sample after illumination with a helium neon laser beam was sensed by means of photodiodes and the digital signals were processed in Lab View software. They also experimented with CMOS line arrays for sensing hairiness. However, the poor cost-effectiveness of the solution may hinder its commercial feasibility. the light beam after passing through a yarn sample containing both polarized and depolarized components which retained the information about yarn hairiness intensity. It was noted that installing the appropriate polarizers in the path of the light beam propagating towards the electronic detector can significantly improve the hair detection by the sensor hence enhancing the overall efficiency of the instrument.

VI. APPLICATION OF IMAGE PROCESSING METHODS

Other than signal and sensor processing there were Various applications of image processing in textiles, ranging from fiber characterization to end product serviceability evaluation, are discussed in a detailed review by Behera. Majumdar reported applications of soft computing techniques in the textile industry and highlighted some developments associated with image processing techniques.

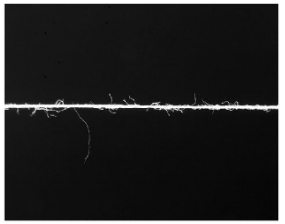


Fig. 2. The original image of yarn.

Utilizing image processing techniques, there are series of research papers about appropriate hardware selection and software development for the measurement of yarn hairiness. Images of yarns are captured by means of a high speed camera in the presence of back lite illumination and yarn transport system. In some research explained about the atomization of capturing. In further research, Algorithm used to process the yarn images already captured. The algorithm was capable of simultaneously identifying the number of hairs and hair length. Many researchers investigated the possibility of application of image processing for the characterization of spun yarns. Firstly the effect of different types of illumination on yarn imaging was examined. It was found that back lit illumination is best for visualizing the structural details of the yarn core. The same authors processed yarn images for hairiness measurement and introduced a new parameter called HDDP (Hairiness density distribution profile), which represents both the number of hairs and their length. They reported an image processing algorithm for yarn hairiness measurement. the focus of their research was to measure the true or intrinsic yarn hairiness including those fibers that are entangled with each other. The algorithm successfully measured the length and number of hairs.

A possible drawback in this method is very small scanning segments of yarns which might not give a true picture of yarn hairiness. Nevertheless the concept of true hairiness measurement is an important one. Researchers from the Technical University of Lodz proposed an algorithm for extraction of the yarn core from a digital image, based on a graph cut method. Chimeh et al. reported an approach for determining yarn hairiness using digital image processing as it plays a vital role in determining bulkiness of hair textured yarns. There are computer algorithm that was able to process the digital images of yarns.

A comparison with microscopic analysis, electrostatic measurement and optical methods showed that their method of hairiness measurement had improved measurement accuracy. Yuvaraj and Nayar proposed a customized method of yarn hairiness measurement, which utilized an image processing algorithm. They also discussed the application of high voltage to erect the hairs from the yarn core during imaging. This technique can help in untangling the hairs from the core and give a better estimate of the hairiness profile of yarn. The commercial feasibility of this method is not yet known. The use of yarn natural images taken from and are processed based on the segmentation, and grey scale specification which is separated by an additive Gaussian noise with different standard deviations.

Incorporating such information will help the algorithm to cope with different image components leading to an adapted filtering window. More explicitly, given the position of an observation, our objective is to know what is the most likely pixel position in the vicinity with similar image content. This could help to define the transitions on the image lattice towards finding similar pixels within a window of a fixed radius. Such a move is by definition anisotropic and is deduced from the observations. For this purpose, the introductions of probability function with respect to the displacement vector.

A. Algorithms

The advanced modeling techniques, i.e. neural networks and fuzzy logics for predicting hairiness, digital image processing, and signal processing for measurement of yarn hairiness and the effects of various material and processing parameters on yarn hairiness using artificial intelligence techniques.

The lists of algorithms are used in the processes of hairiness of the yarn to find the hair index and the parameters required for the process of findings in image processing techniques. Each and every algorithm has its unique style of processing the image for the process of segmentation,

- Kittler binary
- Graph cut Segmentation Algorithm
- Double threshold algorithm
- ICM algorithm
- MRMRF algorithm

VII. EXISTING HAIRINESS MEASUREMENT SYSTEMS LIMITATIONS

A well-established limitation of the hairiness measurement system is the effect of testing speed on yarn hairiness results. Various authors focused on this

issue and discussed the differences in friction and air drag at different test speeds.

Serious limitations of sensor array based instruments reported. Were the effects of sensor resolution, determination of zero reference point on the instrument and signal threshold level on hairiness results. The ability of the instrument to detect hair intersections normally decreases with a decrease in signal threshold level and similarly the determination of the zero reference point affects hairiness measurement. Guha et al. pointed out the effect of orientation of the protruding hair on detection made by commerce all instrument In most cases hairs are not straight or exactly perpendicular to the yarn core but can be Entangled, wrapped around the core, twisted with each other and randomly oriented. This randomness in their structure does not allow conventional instruments to measure their true or intrinsic length. It is worth noting here that even the results of existing systems of hairiness measurement, working on different principles, cannot be correlated. A lot of work remains to be done on harmonization and standardization of hairiness testing. The devices that measure yarn hairiness based on light scattering usually use a stoppage device to block the incident light so the sensor can detect the scattered signal. This stopping phenomenon can cause a loss of information, which may result in wrong measurement of yarn hairiness. All types of existing hairiness measurement systems scan yarn samples in two dimensions for hairiness determination. Only those hairs that spread in front of the sensors will be considered during the measurement. Some hairs may be deflected during testing and only segments of these hairs protrude in front of the sensor. This will result in misclassification of hairs into wrong length groups. This issue certainly needs serious attention from the manufacturer's side. Hairs may be rubbed up during hairiness testing by stationary guides on the testing equipment and produce undesired air drag. Replacing these stationary guides with moveable pulleys could improve the accuracy of hairiness test results. Recently, Haleem and Wang investigated the accuracy of measurement of existing hairiness systems by measuring the actual hairiness of different types of yarns using a tedious yet accurate manual method. They straightened each hair and fixed it to a sheet of paper. The sheet was then scanned and analyzed to obtain true length and number of protruding hairs. The results revealed very significant discrepancies, not just in the number of hairs but also in the actual hair length distribution profile. The hair numbers obtained from this manual method were much greater than that obtained from the hairiness meter, and the true hair length distribution does not follow the well-known exponential decay. Most hairs range in length from Haleem and Wang 1–12mm and then hair frequency decreases gradually up to a length of around 25 mm.

VIII. PREDICTION AND MODELLING YARN HAIRINESS

Advancements in modeling and simulation techniques have allowed quite accurate prediction of yarn properties from fiber and processing parameters. Jackowska-Strumillo *et al.* developed separate artificial neural networks using various quality parameters, including yarn hairiness, for flax/cotton blended spun yarns. Predicted values from the neural model showed good agreement with actual values. It was further observed that finer yarns exhibited less hairiness as compared to coarser ones and the influence of blend ratio cannot be unambiguously determined on yarn hairiness. Similarly, in another study, a comparative analysis was made between regression, non linear regression, perceptron, and ADALINE networks for hairiness prediction. The vital parameters affecting yarn hairiness were yarn linear density and cohesion of fibers in fed sliver. Some other studies also reported hairiness models developed by taking various material and process related parameters as inputs. Both regression analysis and neural networks were utilized for modeling yarn hairiness. However, neural models showed better results while predicting unseen data, pointing towards their ability to model even nonlinear relationship is between variables, which are not possible by line a regression analysis. The logic is also supported by higher values of regression coefficient for neural networks as compared to linear regression.

An important parameter affecting the performance of neural model is the size of training dataset. Comprehensive dataset with appropriate size can train a neural model to predict more efficiently while a limited dataset can significantly decreases its performance. Baykal et al. modeled hairiness of polyester cotton blended yarns by means of regression analysis taking blending ratio and yarn linear density as the input variables. The observe defect of yarn linear density and blend ratio on hairiness agreed the established facts. Finer yarns possessed lower hairiness compared to coarser yarns. Increment in polyester content up to a certain limit in the blend reduced hairiness and later increased hairiness. Later some researchers compared artificial neural network and statistics based regression models for predicting different yarn properties like unevenness and hairiness. An important point here is inclusion of roving parameters in the input variables domain which was not done before. The neural networks performed better than regression analysis here as well indicating the superiority of neural networks due to their ability of understanding nonlinear relations. According to neural model, the decisive parameters that impact yarn hairiness are fiber length uniformity, fiber strength, and its elongation, while regression analysis points out fiber elongation and strength as major affecting parameters. Similarly, Khan *et al.* proposed two hairiness models

based on multiple linear regression and artificial neural networks and compared the results with those obtained from the CSIRO Sirolan Yarn spec program [19]. Neural networks predicted the results closer to target values as compared to regression analysis. Through sensitivity analysis on developed model they found that yarn twist had the greatest effect on hairiness followed by other parameters like ring diameter, average fiber length, fiber diameter, and yarn count. Also, it was observed that relationships between fiber parameter sand yarn properties are not perfectly linear. An interesting publication by Majumdar et al. is worth mentioning here in which the most cost-effective fiber properties are predicted from the desired yarn [20,21]. Comparison between the results of tex cotton rotor yarn from Zweigle hairiness tester and manual method. Properties by reversing the neural network [21]. In this case, yarn hairiness was selected as an input parameter along with other yarn parameters for choosing the appropriate raw material.

The crucial parameters of bale selection, i.e. spinning consistency index and micronaire value of fiber for laydown in blow room, were predicted. The same authors proposed a neuro fuzzy model for predicting cotton yarn hairiness. The analysis showed that two of the most critical parameters for predicting yarn hairiness are yarn count and fiber length. Hairiness decreased when yarn becomes finer and fiber length increased. Mature cotton fibers tended to produce less hairy yarns and short fiber content was not found significantly influencing hairiness. Zhao in his multiple papers reported models of cotton yarn hairiness developed through multi-layer perceptron's. The hairiness of spun yarns was predicted from ring processing parameters, i.e. aperture of guide wire, nip gauge, spindle speed, and back draw time. Also using multi-layer perceptron, the yarn hairiness after sizing and warping processes was predicted through the respective input parameters. However, no information was provided regarding the predictors and their respective influence on yarn hairiness.

IX. CONCLUSIONS

Yarn hairiness is an important quality parameter of Yarns and has vital importance for yarn performance in further processing and for end product performance. Accurate measurement of yarn hairiness is necessary for appropriate quality control. Currently, two hairiness measurement methods based on either a linear array of optical sensors or light scattering principles are commercially in use. Several limitations with these systems raised the necessity of development of new solutions for yarn hairiness determination. Image and signal processing techniques have been investigated as effective and efficient means for evaluating this important yarn characteristic. Successful development of testing hardware and the design of intelligent algorithms has been reported but these solutions have not yet been commercialized. It can be anticipated that novel solutions for hairiness measurement will be available soon because of increasing awareness and demand for quality. Appreciable efforts have been made in modeling yarn hairiness. Various models through statistical and artificial intelligence means were reported predicting the value of yarn hairiness from a wide range of process and material based parameters. The predictions of the models were claimed to be very good. Datasets of acceptable number of observations were utilized to train and validate the neural networks and regression equations. However, neural network based models gave better prediction of output variables than regression equations because of their ability to learn from data patterns and nonlinear fitting of data. Appropriate applications of well-designed models in the spinning industry can serve as a decision support system to achieve high end quality products. In the context of this study and comprehensive review of recent developments in improving yarn hairiness, it is essential to mention some suggestions and possible future work in this area. Measurement of true yarn hairiness is far from reality today because of limitations in existing measurement systems. It is necessary to utilize the intelligent solutions like image and signal processing in combination with promising hardware designs to measure the true hairiness of spun yarns on a commercial scale. Reducing yarn hairiness by means of compact spinning might be an expensive solution for some companies, but simple modifications to existing machines or even to the fibers themselves may also produce yarns with improved quality. In addition to ring spinning, it is important to study other spinning systems with the latest available techniques and technology to further improve them for production of less hairy and highly even yarns.

REFERENCES

[1]. Anirban Guha, C. Amarnath, S. Pateria, and R. Mittal, (2010). 'Measurement of Yarn Hairiness by Digital Image Processing', The Journal of The Textile Institute,Vol.**101**, No.3, pp.214-222.

[2]. Barella A. (1983). Yarn hairiness. Text Prog 1983; **13**: 1–57.

[3]. Barella A. (1993). The hairiness of yarns. Text Prog 1993; **24**: 1–46

[4]. Barella A and Manich AM. (1997). Yarn hairiness update *Text Prog* 1997; **26**: 1–29.

[5]. Barella A and Manich AM. (2002). Yarn hairiness: a further update. *Text Prog* 2002; **31**: 1–44

[6]. Noman Haleem & Xungai Wang (2012). A comparative study on yarn hairiness results from manual test and two commercial hairiness metres School of Textile Science and Engineering, Wuhan Textile University, Wuhan, China Published online: 04 Dec 2012.

[7]. Noman Haleem and Xungai Wang (2014). Recent research and developments on yarn hairiness Textile Research Journal published online 14 July 2014.

[8]. Behera BK. (2004). Image processing in textiles. Text Prog 2004; **35**: 1–193.
[9]. Carvalho V, Soares F and Vasconcelos R. (2009). Artificial intelligence and image processing based techniques: a tool for yarns parameterization and fabrics prediction. In: *IEEE conference on: Emerging technologies and factory automa- tion, Mallorca, 2009.*
[10]. Chimeh MY, Tehran MA, Latifi M, *et al.* (2005). Characterizing bulkiness and hairiness of air-jet textured yarn using imaging techniques. *J Text Inst* 2005; **96**: 251–255.
[11]. Yuvaraj D and Nayar RC. (2012). A simple yarn hairiness measurement setup using image processing techniques. *Indian J Fibre Text Res* 2012; **37**: 331–336.
[12]. Guha A, Amarnath C, Pateria S, *et al.* (2009). Measurement ofyarn hairiness by digital image processing. *J Text Inst 2009*; **101**: 214–222
[13]. Wang X. (1997). The effect of testing speed on the hairiness of ring-spun and sirospun yarns. *J Text Inst* 1997; **88**: 99–106.
[14]. Wang X. (1998). Testing the hairiness of a rotor-spun yarn on the zweigle G565 hairiness meter at different speeds. *J Text Inst* 1998; **89**: 167–169.
[15]. Wang X and Chang L. (1999). An experimental study of the effect of test speed on yarn hairiness. *Text Res J* 1999; **69**: 25–29.
[16]. Wang X, Huang W and Huang X. (1999). Effect of test speed and twist level on the hairiness of worsted yarns. *Text Res. J* 1999; **69**: 889–892.
[17]. Haleem N and Wang X. (2012). A comparative study on yarn hairiness results from manual test and two commercial hairiness meters. J Text Inst 2012: **104**(5): 1–8
[18]. Jackowska-Strumillo L, Jackowski T, Cyniak D, et al. (2004). Neural model of the spinning process for predicting selected properties of flax/cotton yarn blends. *Fibres Text East Eur* 2004; **12**: 17–21.
[19]. Khan Z, Lim AEK, Wang L, et al. (2009). An artificial neuralnetwork-based hairiness prediction model for worsted wool yarns. *Text Res J* 2009; **79**: 714–720.
[20]. Majumdar A, Majumdar PK and Sarkar B. An investigation on yarn engineering using artificial neural networks. *J Text Inst* 2006; **97**: 429–434.
[21]. Majumdar A. (2010). Modeling of cotton yarn hairiness using adaptive neuro-fuzzy inference system. *Indian J FibreText Res* 2010; **35**: 121–127.
[22]. Yuvaraj, D. and Ramesh Chandran Nayar, (2012). 'A Simple Yarn Hairiness Measurement Setup using Image Processing Techniques', *Indian Journal of Fibre & Textile Research*, Vol. **37**, pp.331-336.
[23]. Xiao-Hong Wang, Jing-Yang Wang, Ji-Lei Zhang, Hong-Wei Liang, and Ping-MianKou, (2010). 'Study on the Detection of Yarn Hairiness Morphology *Based on Image Processing Technique', IEEE*, pp.2332-2336.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 08-12(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Survey on Heart Disease: Characteristics, Symptom and Prediction Method

S. Ranilakshmi[1] and Dr. R. Mallika[2]
[1]Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore (Tamilnadu), India.
[2]Assistant Professor, Department of Computer Science, CBM College, Coimbatore (Tamilnadu), India.

ABSTRACT: Advanced data mining techniques are used to discover knowledge in database and for medical research, particularly in Heart disease prediction. In data mining various algorithms techniques are Naïve Bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like K-NN, and Support vector machine (SVM) etc are implemented to find the accuracy of risk factor in heart disease prediction. Several studies have been analyzed to find the risk factor of heart disease which obtains the accuracy by implementing the various data mining algorithms using single technique or comparison techniques of two or more. This paper provides a quick review of available prediction shows the accuracy level of each model using data mining by researchers.

Keywords: Heart Disease, Data Mining Technique, Symptom, Risk Factor, Prediction, Accuracy.

I. INTRODUCTION

"Data Mining is a non-trivial extraction of implicit, previously unknown and potential useful information about data"[1]. In short, it is a process of analyzing data from different perspective and gathering the knowledge from it. The discovered knowledge can be used for different applications for example healthcare industry. Nowadays healthcare industry generates large amount of data about patients, disease diagnosis etc. Data mining provides a set of techniques to discover hidden patterns from data. A major challenge facing Healthcare industry is quality of service. Quality of service implies diagnosing disease correctly & provides effective treatments to patients. Poor diagnosis can lead to disastrous consequences which are unacceptable.

According to survey of WHO, 17 million total global deaths are due to heart attacks and strokes. The deaths due to heart disease in many countries occur due to work overload, mental stress and many other problems. On the whole it is found as primary reason behind death in adults. Diagnosis is complicated and important task that needs to be executed accurately and efficiently. The diagnosis is often made, based on doctor's experience & knowledge. This leads to unwanted results & excessive medical costs of treatments provided to patients. Therefore an automatic medical diagnosis system is designed that take advantage of collected data base and decision.

A. Risk factor of heart disease

A heart attack, also known as a myocardial infarction, occurs when the blood supply to a section of the heart gets blocked. Without the quick restoration of the blood flow, that portion of the heart muscle will die. Heart attacks are most often the result of coronary artery disease (CAD). Cardiovascular disease (CVD) is the leading cause of morbidity and mortality in the world [1,2]. The Emerging Risk Factors Collaboration incorporating 160,309 study participants demonstrated the independent predictive value of C-reactive protein (CRP) for coronary heart disease (CHD) and stroke, the clinical utility of CRP and fibrinogen in CVD risk prediction, and the limited role of adding the novel lipid markers apolipoprotein A-I, lipoprotein (a), apolipoprotein B, and lipoprotein-associated phospholipase A2 to traditional lipid measures [3-5].
A number of factors have been shown that increases the risk of Heart disease [2] [6,7]:

- Family history
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

B. Characteristics of heart disease

Current research on heart disease research has established that it is not a single condition, but refers to any condition in which the heart and blood vessels are injured and do not function properly, resulting in serious and fatal health problems. There are different types of heart diseases, among which the major types are: atherosclerosis, coronary, rheumatic, congenital, myocarditis, angina and arrhythmia.

C. Symptoms of heart disease

Symptoms of the disease differ from person to person. In majority of the cases, there is no early symptom and the disease identifiable only in the advanced stage. Some common symptoms of heart disease are

- chest pain (Angina pectoris);
- strong compressing or flaming sensation in the chest, neck or shoulders;
- discomforts in chest area;
- sweating, light-headedness, dizziness, shortness of breath;
- pain spanning from the chest to arm and neck, and that amplifying with exertion;
- cough;
- palpitations;
- fluid retention

D. Heart disease for people in India

India is seen as the diabetes and coronary heart disease capital of the world. According to current estimates, India will soon have the highest number of heart disease cases in the world. According to the Indian Heart Association, "50% of all heart attacks in Indians occur under 50 years of age and 25% of all heart attacks in Indians occur under 40 years of age. Population living in cities is three times more prone to heart attacks than people living in villages".

Heart disease is increasing in younger generation with a significant risk in both males and females. More and more number of young Indians is suffering from coronary artery disease, owing to their poor lifestyle, and if this continues the future looks even more dangerous. Now, many cases in the 25-35 age groups with heart attack.

Problems faced by young Indians that increases heart risk are:

• No preference given to Health
• No time to exercise
• Stress
• Junk food
• Obesity and smoking.

Preliminary results from the largest study, yet to find out the exact causes of mortality in India, have revealed that heart ailments take most lives in both urban and rural areas. Fig. 1 shows 25% of deaths in the age group of 25-69 years occur because of heart diseases. If all age groups are included, heart diseases account for about 19 % of deaths.

In urban areas, 32.8% deaths occur because of heart ailments, while this percentage in rural areas is 22.9%. It is the leading cause of death among males as well as females. It is also the leading cause of death in all regions. The proportion of deaths caused by heart disease is the highest in south India 25 % and lowest 12% in the central region. This study is being carried out in collaboration with the Registrar General of India (RGI) and the Indian Council of Medical Research (ICMR). About 9.5 million deaths, which are about one in six deaths worldwide, occur in the country every year.

From fig 2 Indians have a 50-400% higher rate of heart disease and diabetes than other populations, regardless of geographic location. Indian women have an equally high disease rate as Indian men.

TOP TEN CAUSES OF DEATH IN INDIA
(Ages 25 to 69 as percentage)

RANK	CAUSE OF DEATH	MALE	FEMALE	TOTAL
1	Cardiovascular diseases	26.3	22.5	24.8
2	Respiratory diseases	10.1	10.4	10.2
3	Tuberculosis	11.4	8.3	10.1
4	Malignant and other tumours	7.8	11.8	9.4
5	Ill-defined conditions	4.8	6.0	5.3
6	Digestive diseases	6.1	3.5	5.1
7	Diarrhoeal diseases	4.0	6.6	5.0
8	Unintentional injuries	5.0	4.1	4.6
9	Intentional self-harm	3.3	2.6	3.0
10	Malaria	2.4	3.4	2.8

(All ages as percentage)

RANK	CAUSE OF DEATH	MALE	FEMALE	TOTAL
1	Cardiovascular diseases	20.3	16.9	18.8
2	Respiratory diseases	9.3	8.0	8.7
3	Diarrhoeal diseases	6.7	9.9	8.1
4	Perinatal conditions	6.4	6.2	6.3
5	Respiratory infections	5.4	7.1	6.2
6	Tuberculosis	7.1	4.7	6.0
7	Malignant and other tumours	5.4	6.0	5.7
8	Senility	4.0	6.5	5.1
9	Unintentional injuries	5.2	4.5	4.9
10	Ill-defined conditions	4.6	5.0	4.8

Fig. 1. Analyzing causes of death in India.

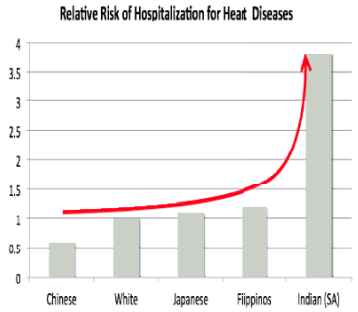


Fig. 2. Comparison in country wise.

II. LITERATURE SURVEY

A review is carried out on different techniques used by researchers in detection of HD. Enormous technologies of DM are involved in design of HD prediction model.

Carlos Ordonez (2004) presented improved study by using rule association for prediction of HD. The study involved finding in detection of HD.

The assessed data set covered medical records of people having heart disease with attributes like chest pain, blood pressure, cholesterol and blood sugar for risk associated factors. The heart per fuser was measured and artery depletion was observed by the author [2].

Yanwei X *et al.* (2007) developed data mining algorithms for predicting survival of Coronary Heart Disease (CHD) patients based on 1000 cases. Because the CHD prediction technique needs a priority for addressing as it was a challenge for medical society. The work carried vast observation on the medical data for 6 months of 1000 CHD records. Information of the survival rate was recorded. The three major DM techniques were employed on 502 cases. "10-fold cross-validation" measures and estimates the accuracy and also performance of these techniques. The measures used were "accuracy, sensitivity and specificity". Confusion matrix was obtained to calculate the three measures. The accuracy obtained was 92.1%, 91.0% and 89.6% for SVM, ANN and DT. A study carried out by comparing different prediction models of CVD patients with a cross validation of 10-fold which gives the glimpse of the ability to handle variety of data. Survival "1" with survivability and non-survival with "0" called binary or Boolean representation was subjected on raw data. Amongst 1000 CHD cases, 798 cases- survival and 202 cases- death [3].

Sellappan Palaniappan *et al.* (2008) proposed an "Intelligent Heart Disease Prediction System (IHDPS)" model designed by applying DT, NB and NN, DM techniques. Result obtained gave strength of each technique which fulfilled the objects. Major queries like "what-if" were answered by IHDPS where general DT could not do. Patterns with relationship type with important knowledge were developed by keeping risk factors of HD. IHDPS model was developed on web, making it more user friendly with scalability vertically as well as horizontally and reliable. Risk factors namely age, blood sugar, BP, family history and other fields were base for detection of HD. 909 information set with 15 attributes collected from"Cleveland" HD database. The DM techniques namely NN,NB and DT were used. Amongst which NB gave best result [4].

Latha *et al.* (2007) used Coactive Neuro Fuzzy Interface Systems and GA in designing HD prediction model which was smart with less error for mean square. After analysis of many methods, two techniques are combined those were NN and GA. Both together form a hybrid method for prediction keeping the account of risk factor and optimize the NN weight. This was the first hybrid method. The primary aim was to use this technique in medical decision support and to indicate the risk for reducing so that it helps patient in reducing the chances of HD further [5].

Carlos Ordonez (2006) solved the limitations of Association rule. Medical record was collected from "Emory University Radiology Department". The algorithm which was designed could search for constraint attributes which intern decreased set of rules. This algorithm looked for association rules on the trained data set and did the validation by testing them independently. Bioinformatics significance of rules found were valued based on support, confidence. Again it was checked and evaluated using lift. The data consisted of records of the patients with HD. Searching, training and testing reduced association rules and accuracy rate was high [6].

K. Srinivas *et al.* (2010) analyzed the "behavioral Risk Factor Surveillance System" where it was surveyed by testing and giving a self-report on CVD rate by telling that rate of CVD is high in coal mining region by carrying out tests on Singareni Collieries Company in Andhra Pradesh, India, compared to other regions after control for other risks. Diagnosis was done by collecting patient's report which gave records of distinguished measures saying about the morbidity. Along with these attributes the data sets which are symptoms of HD were also considered. Sensitivity and accuracy were the two key measures for evaluation. (Accuracy and sensitivity) obtained was (0.825% & 0.88) for SVM, (0.825% & 0.8717) for DT C4.5 and (0.897%, 0.9017) for NN- Multi layered per ceptron [8].

HeonGyu Lee *et al.* (2007) used multiparametric features like linear and non-linear features of high-rate variability (HRV) of 3 posture positions namely supine, left lateral and right lateral position to find the HRV indices to detect coronary artery disease (CAD). Bayesian classification, associative classifier, classification based on multiple association rule (CMAR), C4.5 (DT) and SVM classifiers are used for predicting coronary artery diseases. Statistical analysis were used for essential feature selection. Accuracy of SVM, CMAR, C4.5, NB (Tree Augmented NB- TAN) and NB (Selection Tree Augmented NB- STAN) were 90%, 80%, 78%, 81% and 85% respectively. SVM showed the best performance [9].

Hongmei Yan *et al.* (2006) used Multilayer perception (MLP) with 40 input variables for Input layer and the output layer with 5 nodes. Improved back propogation algorithm was used to train the system. 352 medical records were collected for train & test the system. Assessment methods like cross validation, holdout and bootstrapping were applied to assess the system. MLP is well known for its good architecture which makes this method most important in NN models also its algorithms are very easy and simple in understanding. MLP consists of 3 layers namely Input layer, Hidden Layer and Output Layer.

The hidden layer was obtained by cascading learning process. The experimental results achieved were of 90% accuracy [10].

Shantakumar B. Patil *et al.* (2009)"IEHPS, heart attack prediction system" proposed was a smart & efficient technique designed using NN. Significant frequent patterns are generated "by 'K-means clustering' and 'MAFIA' algorithm". [11]

Peter, T.J. *et al.* (2012) used classification data mining for prediction of heart disease. Input set has intrinsic linear combination of variables which cannot be adapted at modeling nonlinear complex interaction in medical domains. These limitations of conventional medical scoring systems are handled by using classification models. The data is cleaned by using classification method of DM where the complex and implicit relationship between interdependent variables is detected. Naïve bayes, k-NN, decision tree and neural network were implemented on the knowledge data. Naïve bayes performed better than other methods. Accuracy of the classifiers was 83.70%, 76.66%, 75.18% and 78.148% of NB, DT, K-NN, NN respectively [12].

Mai Shouman *et al.* (2012) proposed a model by using single data mining technique and hybrid data mining technique. The kernel density, automatically defined groups, bagging algorithm and support vector machine were focused. The accuracy generated was 84.1% [13].

Syed Umar *et al.* (2013) presented a system by combining two techniques called as hybrid model by using global optimization benefits of GA specifically to initiate the NN weights. Back propogation algorithm was used for learning and training the neural network. A multi-layered feed forward network had "12-input, 10 hidden and 2 output nodes". Input count depends on final risk factor of the chronic. Training and learning is done by "The Levenberg Marquardt back propagation algorithm". Bias and weight is recorded and updated by network training function. Implementation of the system was done by using "Matlab R2012a, Global Optimization Toolbox and the Neural Network Toolbox". Risk factors of 50 patients was collected and the results obtained showed training accuracy of 96.2% and a validation accuracy of 89% [14].

Shamsher Bahadur Patel *et al.* (2013) deduced 14 attributes to 6 by using GA. Then Classifiers NB, classification by clustering & DT were used to predict the diagnosis of HD. Genetic search was applied on the 14 attributes and no. of attributes was reduced to 6. The 6 attributes were Resting BP (RBP), old peak (OLDPK), chest pain type (CPTYPE), number of major vessel colored (VSL), Exercise induced angina (EIA) and max heart rate achieved (THAL). This reduced dataset was applied to 3 classification models. For designing the model 4 computing evaluation measures used were namely, True positives (T-POS) refers to positive tuple, True Negative (T-Neg) refers to negative tuple, False Negative (F-Neg) and False positive (F-POS). WEKA tool was used for implementation. Accuracy obtained by DT, NB, Classclust was 99.2%, 96.5% and 88.3% respectively. It was observed that DT DMT outperformed the other two DM technique, WEKA – Waikato Environment for Knowledge analysis [15].

I.S. Jenzi *et al.* (2013) designed a reliable classifier model using data mining technique, Association rules and classification techniques like decision tree, Naïve Bayes and Neural Network. The relationship between key patterns was established. The data set consists of 14 attributes, where class attribute is considered at the end of all attributes. The GUI was developed in Microsoft .NET platform, with interconnections done by using IKVM interface with some Java libraries. The receiver operating characteristic (ROC) curves was a visual tool used to represent the accuracy. The results obtained was represented on ROC which has the area under ROC of data mining was 0.807 which is found better than Naïve Bayes [16].

G. Purusothama *et al.* (2015) compared different classification techniques to design risk prediction model for HD. The two kinds of models compared were 1. Single model to test data 2.Combined model to test data (hybrid model). These two models were used for data analysis. The authors in this survey have considered only the classification techniques in case of single model and combined model. Accuracy obtained for decision technique, association rule, K-NN, artificial neural network, naive bayes, hybrid approach were 76%, 55%, 58%, 85%, 86%, 69% and 96% respectively. It was concluded that by applying hybrid data mining techniques promising results were obtained in diagnosis of heart disease [17].

Minas A. Karaolis *et al.* (2009) used "Event-related risk factors" investigated for 1. Before the event: non-modifiable and modifiable (risk factors were family history, hypertension, age, sex, smoking and blood sugar). 2. After the event: modifiable-smoking, systolic/diastolic blood pressure, cholesterol. 528 patients records collected from "Paphos district in Cyprus". "Myocardial infarction (MI), percutaneous coronary intervention (PCI) and coronary artery bypass graft surgery (CABG)" were the events used for investigation. The accuracy obtained were 66%, 75%, and 75% for MI, PCI and CABG models respectively [18].

Hlaudi Daniel Masethe *et al.* (2014) proposed DM algorithm namely J48, NB, "REPTREE CART" and Bayes Net were used for predicting "heart attacks". The data set was collected from hospital and doctors who were practitioners in South Africa. 11 attributes considered were patient Identification Number, Gender, age, chest pain, cardiogram, BP, rating of heartbeat, cholesterol, tobacco 259 consumption, hot drinks intake & diabetes level. The tool called Waikote Environment for knowledge Analysis (WEKA) was used for prediction of HD. WEKA tool was significant in discovering analyzing and predicting patterns. Accuracy obtained were 99.0741, 99.222, 98.148, 99.0741 for J48, REPTREE, Naïve Bayes, Bayes Net and simple CART respectively, Bayes Net algorithm outperformed the N B algorithm [19].

Lokanath Sarangi *et al.* (2015) designed a cost efficient model by using Genetic Algorithm optimizer technique. The weights were optimized and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks [20].

M. Satish, *et al.* (2015) used different DM techniques (DMT) like Rule based, DT, NB, and ANN. An efficient approach called pruning-classification association rule (PCAR) was used to generate association rules from HD warehouse for prediction of Heart Disease. Pima Indian Heart attack data warehouse was used for pre-processing for mining. All the above mentioned DMT were explained [21].

Chaitrali S. *et al.* (2012) used 13 attributes like sex, blood pressure, and cholesterol for prediction of heart disease. Two more attributes called smoking and obesity was added. DM classification methods used were NN, DT and NB. Accuracy obtained for these techniques were 100%, 99.62% and 90.74% respectively. Confusion matrix was obtained for 3 classification methods for 13 attribute data sets and 15attribute data set. The accuracy with 15 attribute was 100% for neural network [22].

III. CONCLUSION

Cardiovascular risk factors are the leading causes of death worldwide. The identification of risk factors provides new opportunities for developing more effective approaches for preventing and treatment of cardiovascular disease.

REFERENCES

[1]. World Health Organization. World Health statistics Annual, Geneva, Switzerland: World Health Organization (2006), available from: http://www.who.int/ mediacentre/ factsheets/ fs310.pdf.

[2]. Carlos Ordonez. Improving heart disease prediction using constrained association rules. Presented in a Seminar at University of Tokyo; 2004.

[3]. Yanwei X, Wang J, Zhao Z, Gao Y. Combination data mining models with new medical data to predict outcome of coronary heart disease. *Proceedings International Conference on Convergence Information Technology;* 2007. p. 868–72.

[4]. Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS *International Journal of Computer Science and Network Security*, Vol. **8** No.8, August 2008.

[5]. Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", *International Journal of Biological and Life Sciences*, Vol.**3**, No.3, pp.157-160, 2007.

[6]. Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction, Carlos Ordonez *IEEE Transactions On Information Technology In Biomedicine, VOL. 10, NO. 2, APRIL 2006.*

[7]. Personal Heart Monitoring and Rehabilitation System using Smart Phones Peter Leijdekkers, Valérie Gay, Proceedings of the International Conference on Mobile Business (ICMB'06) 0-7695-2595-4/06 $20.00 © 2006 IEEE.

[8]. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques, K. Srinivas,G. Raghavendra Rao ; A. Govardhan, in *IEEE 5th International Conference on Computer Science and Education (ICCSE), 2010.*

[9]. Heon Gyu Lee, Ki yong Noh and Keun Ho Ryu, " Mining Biosignal Data: Coronary Artery Disease diagnosis Using Linear and Nonlinear Features of HRV", Springer-Verlag Berlin Heidelberg 2007.

[10]. Hongmei Yan, Yingtao Jiang, Jun Zheng, Chenglin Peng and Qinghui Li, " A Multilayer perceptron-based medical decision support system for heart disease diagnosis", ELSEVIER 2006.

[11]. Shantakumar B. Patil and Y.S. Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European *Journal of Scientific Research*, Vol. **31**, No.4, pp.642-656, 2009.

[12]. An empirical study on prediction of heart disease using classification data mining techniques, Peter, T.J, . Somasundaram, K., 2012 *International Conference on Advances in Engineering, Science and Management (ICAESM).*

[13]. Mai Shouman, Tim Turner, Rob Stocker, " Using data mining techniques in heart disease diagnosis and treatment", *IEEE Japan-Egypt Conference on Electronics, Communications and Computers, 2012.*

[14]. Genetic Neural Network based Data mining in prediction of Heart disease using risk factors. Syed Umar Amin1, Kavita Agarwal2, Dr. Rizwan Beg *Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).*

[15]. Shamsher Bahadur Patel, Pramod Kumar Yadav and Dr. D.P. Shukla, "Predict the Diagnosis of Heart Disease Patients using classification Mining Techniques", *IOSR Journal of Agriculture and Veterinary Science* (IOSR-JAVS), 2013.

[16]. I.S. Jenzi, P. Priyanka, Dr. P. Alli, "A Reliable Classifier Model Using Data Mining Approach for Heart Disease Prediction", *International Journal of Advanced Research in* Computer Science and Software Engineering, 2013.

[17]. G. Purusothama and P. Krishnakumari, "A Survey of Data mining techniques on risk prediction: Heart disease", *Indian Journal of Science and Technology, 2015.*

[18]. Minas A. Karaolis, Joseph A. moutiris, Dementra Hadjipanayi, " Assessment of the risk factors of coronary heart events based on data mining with decision trees", *IEEE transactions on information technology in biomedicine, 2010.*

[19]. Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using classification Algorithm", *Proceedings of the World Congress on Engineering & Computer Science 2014, WCECS.*

[20]. Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", IJCTA, International Press 2015.

[21]. M.Satish, D Sridhar, "Prediction of Heart Disease in Data Mining Technique", *International Journal of Computer Trends & Technology (IJCTT), 2015.*

[22]. Chaitrali S. Dangare Sulabha S Apte, "Improve study of Heart Disease prediction system using Data Mining Classification techniques", *International journal of computer application, 2012.*

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 13-18(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Survey on Cloud Service Security Schemes using Cryptography Techniques

T. Rajeshwari[1] and C. Thangamani[2]
[1]Research Scholar, P.K.R Arts College for Women Gobichettipalayam (Tamilnadu), India.
[2]Associate Professor, P.K.R Arts College for Women Gobichettipalayam (Tamilnadu), India.

ABSTRACT: The cloud data centers are used to provide shared data for the users. The cloud services are protected with the cryptography and digital signature methods. The hybrid cryptography models are applied for the cloud service security operations. The Hybrid Cryptography technique combines the Advanced Encryption Standard (AES), Blowfish, Rivest Cipher 6 (RC6) and Byte rotation algorithm (BRA) based symmetric cryptography methods. The encrypted file is transferred to the cloud user with reference to the request value. The cloud services are secured with Collaborative Cryptography and digital signature techniques. The symmetric and asymmetric cryptography algorithms are integrated in the Collaborative cryptography model. The RSA and Elliptic Curve Cryptography (ECC) algorithms are applied to improve the security levels. The data hiding methods are adapted to support the key distribution tasks. The Message Digest 6 (MD6) Algorithm is adapted for the data integrity verification process.

Keywords: Cloud Data Centers, Cloud service security, Symmetric cryptography, Asymmetric cryptography and Digital signatures

I. INTRODUCTION

Cloud Computing is internet based technology which has evolved in the field of IT over the past few years. Cloud computing makes the transfer or storage of bulk data easy to be transferred and maintained for usage. Organizations need not buy special hardware for deploying different applications since cloud computing provides with pay-as-you-go pricing basis which means that all the resources like firewall, server, database and so on that are required by an organization for the deployment of an application may be leased out by some other organization which deals in providing those resources. The latter organizations are known as cloud vendors. Hence leasing out of resources does not levy high cost on the users and at the same time it gives business to other people as well. So, cloud computing is becoming popular to in the field of IT and is gaining attention of various organizations.

Security is one of the most difficult task to implement in cloud computing. The paper basically deals with the security issues that are experienced during the storage of data on the cloud. The cloud vendors generally store the client's data and information in cloud without following any security measures. Almost every cloud provider does not provide enough security measures to ensure the data safety and that's why clients waver keeping their data at some place which is very easy to be accessed by someone else [5].

II. CLOUD SECURITY USING HYBRID CRYPTOGRAPHY

Cryptography technique translates original data into unreadable form. Cryptography technique is divided into symmetric key cryptography and public key cryptography. This technique uses keys to translate data into unreadable form. So only authorized person can access data from cloud server.

Symmetric key cryptography algorithms are AES, DES, 3DES, IDEA, BRA and blowfish. The main issue is to deliver the key to receiver into multi user application. These algorithm require low delay for data encode and decode but provides low security. Public key cryptography algorithm is RSA and ECC algorithm. Public and private keys are manipulated into public key cryptography algorithms. These algorithms accomplished high level security but increase delay for data encode and decode. Steganography hide the secret data existence into envelope. In this technique existence of data is not visible to all people. Only valid receiver knows about the data existence. Text steganography technique is used to produce high security for data. Secret data of user hide into text cover file. After adding text into text cover file it looks like normal text file. If text file found by illegitimate user than also cannot get sensitive data. Advantage of text steganography technique is provide security to text. Minimum space is essential for text steganography as compare to image steganography [1].

Three bit LSB technique used for image steganography [3]. This system is suggested by author R.T. Patil. Sensitive data of user hide into cover image. We can hide huge amount of into image using LSB steganography technique .The author Klaus Hafmann has implemented high throughput architecture for cryptography algorithm. AES is symmetric key cryptography algorithm. It supports three types of keys. For 128 bit key require 10 rounds, [1]192 bit key require 12 rounds and 256 bit key require 14 rounds. In improved AES algorithm encryption and decryption time is reduced .Advantage of modified AES algorithm is provides better performance in terms of delay.

New symmetric key cryptography algorithm is presented by author M. Nagle. It applies a single key for texts encode and decode. Size of key is 128 bit. Many steps are executed randomly so illegitimate user can even guess the steps of algorithm. Providing high throughput is one of the advantages of symmetric key cryptography algorithms. Improved DES algorithm uses 112 bit key size for data encode and decode. For data encode purpose two keys are used.128 bit input of DES algorithm is divided into two parts. That two parts are executed at a same time. DES algorithm has one weakness. That is less key size. 3DES algorithm essential large amount of time for encryption and decryption. Improved DES algorithm has capability of provide better performance as compare to DES and 3DES. Name Based Encryption Algorithm is work on one byte at a time. It uses secret key for encryption and decryption. Key generation process is done using random key generation technique. It provides security to data. Disadvantage of this algorithm is essential maximum time for converting data into cipher text because it operate on single byte at a time. To solve data storage and security issues author has new security model .In this model private and public cloud storage areas are used for increase security level of data. On private cloud secure data is stored and unnecessary data is stored on public cloud. The main reason behind this system is reduce storage cost .Private cloud is more secure than the public cloud. To enhance security of file in cloud computing, source file is break into different into different part. Every part of file is encrypted and stored on more than one cloud. Information about file is stored on cloud server for decryption purpose. If attacker tries to recover original file than t he will get only a single part of file. Elliptic Curve cryptography algorithm is used to accomplish high level security. Key managing complications are removed using access management and identity.ECC algorithm need maximum amount of time for file encode and decode. File is converted into unreadable format using AES algorithm. Encrypted file is stored on cloud.AES algorithm is less secure than public key cryptography algorithms [2].

AES and 3DES algorithms are merged into hybrid algorithm to accomplish confidentiality. It is harder for attacker to recover secret file of user. It consumes maximum amount of delay to translate data into decode and encode form. In existing system single algorithm is used for data encode and decode purpose. But use of single algorithm is not accomplish high level security. If we use single symmetric key cryptography algorithm than we have to face security problem because in this type of algorithm applies a single key for data encode and decode.

So key transmission problem occur while sharing key into multiuser environment. Public key cryptography algorithms accomplish high security but maximum delay is needed for data encode and decode. To solve above issues we have introduced new security mechanism [5].

Cloud owner and cloud user are included into system architecture. Cloud owner upload the data on cloud server. File is split into octet. Every part of file is encoded simultaneously using multithreading technique. Encoded file is stored on cloud server. Keys used for encryption are stored into cover image. Cloud computing is the multi user environment .In this more than one user can access file from cloud server. Cloud user request for file. On request of file user also get stego image using email which consist of key information. Reverse process is used for decode the file [3].

III. PROBLEM STATEMENT

The cloud file storage security is provided with Hybrid Cryptography technique. The encryption/decryption operations are performed with 128 bits key based symmetric cryptography methods. The Hybrid Cryptography technique integrates the Advanced Encryption Standard (AES) and Blowfish algorithm. The Rivest Cipher 6 (RC6) and Byte rotation algorithm (BRA) are also combined in the Hybrid Algorithm. The data file is divided into 8 segments and each segment is encrypted with a secret key cryptography algorithm. The encrypted file is transferred to the cloud user with reference to the request value. The image steganography is employed to transfer the key values to the cloud users. The Least Significant Bit (LSB) encoding scheme is used for the image steganography. The following issues are identified from the current encrypted cloud services.

IV. SYMMETRIC AND ASYMMETRIC KEY CRYPTOGRAPHY MODELS

The data security is ensured with cryptography techniques. The symmetric and asymmetric cryptography techniques are applied for the data transmission security process in the cloud services. The symmetric cryptography model uses the same key for the encryption and decryption process. The asymmetric cryptography odel uses the public key for the data encryption and private key for the decryption process.

The symmetric cryptography algorithms are build with single key model. The same key is used for the encryption and decryption process. All the encryption and decryption operations are carried out with permutation and computation methods. Advanced Encoding Standard (AES), Blowfish, Byte – Rotation Encryption and Rivest cipher 6 Algorithm are most popularly used symmetric key cryptography algorithms for all types of data values.

A. Advanced Encoding Standard (AES)

This standard specifies the Rijndael algorithm a symmetric block cipher that can process data blocks of 128 bits, using cipher keyswith lengths of 128, 192, and 256 bits. Rijndael was designed to handle additional block sizes and key lengths, however they are not adopted in this standard [2].

Throughout the remainder of this standard, the algorithm specified herein will be referred to as the AES algorithm. The algorithm may be used with the three different key lengths indicated above, and therefore these different "flavors" may be referred to as "AES-128", "AES-192", and "AES-256". This specification includes the following sections:

1. Definitions of terms, acronyms, and algorithm parameters, symbols, and functions;
2. Notation and conventions used in the algorithm specification, including the ordering and numbering of bits, bytes, and words;
3. Mathematical properties that are useful in understanding the algorithm;
4. Algorithm specification, covering the key expansion, encryption, and decryption routines;
5. Implementation issues, such as key length support, keying restrictions, and additional block/key/round sizes.

B. Blowfish Algorithm

Blowfish is a symmetric-key block cipher, designed in 1993 by Bruce Schneier and included in a large number of cipher suites and encryption products. Blowfish provides a good encryption rate in software and no effective cryptanalysis of it has been found to date. However, the Advanced Encryption Standard (AES) now receives more attention.

[7] Schneier designed Blowfish as a general-purpose algorithm, intended as an alternative to the aging DES and free of the problems and constraints associated with other algorithms. At the time Blowfish was released, many other designs were proprietary, encumbered by patents or were commercial or government secrets. Schneier has stated that, "Blowfish is unpatented, and will remain so in all countries. The algorithm is hereby placed in the public domain, and can be freely used by anyone. Notable features of the design include key-dependent S-boxes and a highly complex key schedule.

Blowfish has a 64-bit block size and a variable key length from 32 bits up to 448 bits. It is a 16-round Feistel cipher and uses large key-dependent S-boxes. In structure it resembles CAST-128, which uses fixed S-boxes. Each line represents 32 bits. There are five subkey-arrays: one 18-entry P-array (denoted as K in the diagram, to avoid confusion with the Plaintext) and four 256-entry S-boxes (S0, S1, S2 and S3).

Every round *r* consists of 4 actions: First, XOR the left half (L) of the data with the *r* th P-array entry, second, use the XORed data as input for Blowfish's F-function, third, XOR the F-function's output with the right half (R) of the data, and last, swap L and R.

The F-function splits the 32-bit input into four eight-bit quarters, and uses the quarters as input to the S-boxes. The S-boxes accept 8-bit input and produce 32-bit output. The outputs are added modulo 2^{32} and XORed to produce the final 32-bit output (see image in the upper right corner). After the 16th round, undo the last swap, and XOR L with K18 and R with K17 (output whitening).

Decryption is exactly the same as encryption, except that P1, P2, …, P18 are used in the reverse order. This is not so obvious because xor is commutative and associative. A common misconception is to use inverse order of encryption as decryption algorithm (i.e. first XORing P17 and P18 to the ciphertext block, then using the P-entries in reverse order).

Blowfish's key schedule starts by initializing the P-array and S-boxes with values derived from the hexadecimal digits of pi, which contain no obvious pattern (seenothing up my sleeve number). The secret key is then, byte by byte, cycling the key if necessary, XORed with all the P-entries in order. A 64-bit all-zero block is then encrypted with the algorithm as it stands. The resultant ciphertext replaces P_1 and P_2. The same ciphertext is then encrypted again with the new subkeys, and the new ciphertext replaces P_3 and P_4. This continues, replacing the entire P-array and all the S-box entries. In all, the Blowfish encryption algorithm will run 521 times to generate all the subkeys - about 4KB of data is processed[6].

Because the P-array is 576 bits long, and the key bytes are XORed through all these 576 bits during the initialization, many implementations support key sizes up to 576 bits. While this is certainly possible, the 448 bits limit is here to ensure that every bit of every subkey depends on every bit of the key as the last four values of the P-array don't affect every bit of the ciphertext. This point should be taken in consideration for implementations with a different number of rounds, as even though it increases security against an exhaustive attack, it weakens the security guaranteed by the algorithm. And given the slow initialization of the cipher with each change of key, it is granted a natural protection against brute-force attacks, which doesn't really justify key sizes longer than 448 bits.

C. Byte – Rotation Encryption Algorithm

The "Byte – Rotation Encryption Algorithm (BREA)" is applied on different blocks of plaintext and executes in parallel manner through multithreading concept of single processor system [4]. This paper is an

attempt to invent a new encryption model which is more secure and very fast to others.

The steps of proposed Byte-Rotation Encryption Algorithm:

1. The letters of alphabet are assigned numerical values from 1 to 26 in sequence i.e. A, B, C, ..., X, Y, Z assigned numerical values 1, 2, 3, ..., 24, 25, 26 respectively, the digits from 1 to 9 assigned numerical values from 27 to 35 respectively and the zero (0) remains as it is.

2. The plaintext is partitioned into fixed-length blocks of size 16 bytes (or 128 bits) each. These blocks are represented by a matrix Mp.

3. The values of Key matrix (K) are randomly selected from the range 1 to 26. The size of Key matrix is equivalent to the block size of plaintext i.e. 16 bytes. K = [k1, k2, ..., k16] K = Random (1, 26, 16)

4. Calculate the Transpose matrix of plaintext block matrix (Mp), which is denoted by MpT.

5. Calculate encrypted Key matrix Ke using the following formula: Ke = K mod 2

6. Add both the matrices MpT and Ke and the resultant matrix is denoted by Cpk. Cpk = MpT + Ke

7. Rotate first three rows horizontally of Cpk matrix such that rotate one byte from first row, rotate two bytes from second row, rotate three bytes from third row and fourth row remains untouched. The resultant matrix is denoted by Chr.

8. Rotate first three columns vertically of Chr matrix such that rotate one byte from first column, rotate two bytes from second column, rotate three bytes from third column and fourth column remains untouched. The resultant matrix is denoted by Cvr.

9. Replace numeric values of Cvr matrix by their corresponding letters and if 36 exist in Cvr matrix, it is replaced by the special character #. The resultant matrix is denoted by Ce.

D. Rivest cipher 6 Algorithm

In cryptography, **RC6** (**Rivest cipher 6**) is a symmetric key block cipher derived from RC5. It was designed byRon Rivest, Matt Robshaw, Ray Sidney, and Yiqun Lisa Yin to meet the requirements of the Advanced Encryption Standard (AES) competition. The algorithm was one of the five finalists, and also was submitted to the NESSIE and CRYPTREC projects. It was a proprietary algorithm, patented by RSA Security.

RC6 proper has a block size of 128 bits and supports key sizes of 128, 192, and 256 bits up to 2040-bits, but, like RC5, it may be parameterised to support a wide variety of word-lengths, key sizes, and number of rounds. RC6 is very similar to RC5 in structure, using data-dependent rotations, modular addition, and XOR operations; in fact, RC6 could be viewed as interweaving two parallel RC5 encryption processes, although RC6 does use an extra multiplication operation not present in RC5 in order to make the rotation dependent on every bit in a word, and not just the least significant few bits [6].

E. Asymmetric Key Algorithms

The asymmetric key cryptography algorithms uses the key pairs for the security operations. The public key is used for the encryption process. The private key is used for the decryption process. The RSA and Elliptic Curve Cryptography (ECC) are the most popularly used asymmetric key cryptography algorithms.

RSA Algorithm. The domain name service sensitive attributes are secured using the RSA algorithm. The Rivert, Shamir, Adelman (RSA) scheme is a block cipher in which the Plaintext and cipher text are integers between 0 and n-1 for some n. A typical size for n is 1024 bits or 309 decimal digits.

Key Generation

Select p,q p and q both prime , $p \neq q$

Calculate $n = p \times q$

Calculate $\phi(n)=(p-1)(q-1)$

Select integer e $\gcd(\phi(n),e) = 1; 1 < e < \phi(n)$

Calculate d $d = e^{-1} \bmod \phi(n)$

Public key $KU = \{e, n\}$

Private key $KR = \{d, n\}$

Encryption

PlaintextM <n

Cipher text $C = M^e \pmod n$

Decryption

Cipher text C

PlaintextM = C^d (mod n)

Elliptic curve cryptography (**ECC**). Elliptic curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. ECC requires smaller keys compared to non-ECC cryptography (based on plain Galois fields) to provide equivalent security.

Elliptic curves are applicable for key agreement, digital signatures, pseudo-random generators and other tasks. Indirectly, they can be used for encryption by combining the key agreement with a symmetric encryption scheme. They are also used in several integer factorization algorithms based on elliptic curves that have applications in cryptography, such as Lenstra elliptic curve factorization.

Public-key cryptography is based on the intractability of certain mathematical problems. Early public-key systems are secure assuming that it is difficult to factor a large integer composed of two or more large prime factors. For elliptic-curve-based protocols, it is assumed that finding the discrete logarithm of a random elliptic curve element with respect to a publicly known base point is infeasible: this is the "elliptic curve discrete logarithm problem" (ECDLP). The security of elliptic curve cryptography

depends on the ability to compute a point multiplication and the inability to compute the multiplicand given the original and product points. The size of the elliptic curve determines the difficulty of the problem.

The primary benefit promised by elliptic curve cryptography is a smaller key size, reducing storage and transmission requirements, i.e. that an elliptic curve group could provide the same level of security afforded by an RSA-based system with a large modulus and correspondingly larger key: for example, a 256-bit elliptic curve public key should provide comparable security to a 3072-bit RSA public key [8].

The U.S. National Institute of Standards and Technology (NIST) has endorsed elliptic curve cryptography in its Suite B set of recommended algorithms, specifically elliptic curve Diffie–Hellman (ECDH) for key exchange and Elliptic Curve Digital Signature Algorithm (ECDSA) for digital signature. The U.S. National Security Agency (NSA) allows their use for protecting information classified up to top secret with 384-bit keys. However, in August 2015, the NSA announced that it plans to replace Suite B with a new cipher suite due to concerns about quantum computing attacks on ECC. While the RSA patent expired in 2000, there may be patents in force covering certain aspects of ECC technology. However some argue that the US governmentelliptic curve digital signature standard (ECDSA; NIST FIPS 186-3) and certain practical ECC-based key exchange schemes.

V. MESSAGE-DIGEST (MD6) ALGORITHM

The MD6 Message-Digest Algorithm is a cryptographic hash function. It uses a Merkle tree-like structure to allow for immense parallel computation of hashes for very long inputs. Authors claim a performance of 28 cycles per byte for MD6-256 on an Intel Core 2 Duo and provable resistance against differential cryptanalysis. The source code of the reference implementation was released under MIT license.

Speeds in excess of 1 GB/s have been reported to be possible for long messages on 16-core CPU architecture. The design of Merkle tree is based on the claims from Intel describing the future of hardware processors with tens and thousands of cores instead of the conventional uni-core systems. With this in mind, Merkle tree hash structures exploit full potential of such hardware while being appropriate for current uni/dual core architectures.

In December 2008, Douglas Held of Fortify Software discovered a buffer overflow in the original MD6 hash algorithm's reference implementation. This error was later made public by Ron Rivest on 19 February 2009, with a release of a corrected reference implementation in advance of the Fortify Report [9].

MD6 was submitted to the NIST SHA-3 competition. However, on July 1, 2009, Rivest posted a comment at NIST that MD6 is not yet ready to be a candidate for SHA-3 because of speed issues, a "gap in the proof that the submitted version of MD6 is resistant to differential attacks", and an inability to supply such a proof for a faster reduced-round version, although Rivest also stated at the MD6 website that it is not withdrawn formally. MD6 did not advance to the second round of the SHA-3 competition. In September 2011, a paper presenting an improved proof that MD6 and faster reduced-round versions are resistant to differential attacks was posted to the MD6 website.

VI. COLLABORATIVE CRYPTOGRAPHY SCHEME

The cloud data security system is build to protect the shared data files maintained under the cloud data centers. The data owner or data provider uploads the data files for the shared access purpose. The Hybrid Cryptography scheme combines the symmetric key cryptography algorithms. The Blowfish, Byte Rotation Algorithm (BRA), Rivest Cipher (RC6) and Advanced Encryption Standard (AES) algorithms are used in the integrated manner. The algorithms are randomly applied for separate blocks in the same files. The Collaborative Cryptography scheme combines both symmetric key cryptography and asymmetric key cryptography technique. The hybrid cryptography scheme is enhanced with RSA and Elliptic Curve (ECC) algorithms. The authentication and integrity verification operations are carried out with the support of the Message Digest (MD6) algorithm. The key values are maintained under the key distribution center. The key values are transferred and updated in the image files. The data hiding techniques are used to protect the key values. The Least Significant Bit (LSB) encoding scheme is used for the data hiding process. The user can download file from the cloud data center and extract the key from the key distribution with block information. The data decryption and verification operations are carried out under the client side.

VII. CONCLUSION

The encrypted cloud services are build to support secured data sharing operations. The data owner maintains the shared files under the cloud server for the cloud users. The hybrid cryptography scheme combines the symmetric cryptography algorithms with Steganography based key exchange models. The symmetric and asymmetric cryptography algorithms are integrated in the Collaborative Cryptography technique with digital signatures. The collaborative cryptography scheme improves the data security for the encrypted cloud services. The Key Distribution Center (KDC) manages the key transmission based on user requests. Data transmission loses are detected with the support of digital signatures. Data security operations are carried out with multi threaded model.

REFERENCES

[1]. Aarthi Singh, Manisha Malhotra, "Hybrid two-tier framework for improved security in cloud environment", IEEE Oct 2016.

[2]. Abu Marjan, Palash Uddin, "Developing Efficient Solution to Information Hiding through text steganography along with cryptography", *IEEE*, *IFOST*, pages 14-17, October 2014.

[3]. Bin Li, Ming Wang, Xiaolong Li, Shunquan Tan and Jiwu Huang, "A Strategy of Clustering Modification Directions in Spatial Image Steganography", *IEEE Transactions on Information Forensics And Security*, Vol. **10**, No. 9, September 2015

[4]. Bingwen Feng, Wei Lu and Wei Sun, "Secure Binary Image Steganography Based on Minimizing the Distortion on the Texture", IEEE Transactions On Information Forensics And Security, Vol. 10, No. 2, February 2015

[5]. Cosimo Anglano, Rossano Gaeta and Marco Grangetto, "Securing coding-based cloud storage against pollution attacks", *IEEE Transactions on Parallel and Distributed Systems*, May 2017.

[6]. Helei Cui, Xingliang Yuan and Cong Wang, "Harnessing Encrypted Data in Cloud for Secure and Efficient Mobile Image Sharing",- *EEE Transactions on Mobile Computing,May 2017.*

[7]. Kaiping Xue, Shaohua Li and Peilin Hong, "Two-Cloud Secure Database for Numeric-Related SQL Range Queries With Privacy Preserving", *IEEE Transactions on Information Forensics and Security* July 2017.

[8]. Nirnay Ghosh, Debangshu Chatterjee, Soumya K Ghosh and Sajal K Das, "Securing Loosely-coupled Collaboration in Cloud Environment through Dynamic Detection and Removal of Access Conflicts", *IEEE Transactions on Cloud Computing*, June 2016.

[9]. Shreyank N Gowda, "Using Blowfish encryption to enhance security feature of an image", 6th International Conference on Information Communication and Management Dec 2016.

[10]. Shulan Wang, Joseph K. Liu, Jianping Yu, Jianyong Chen and Weixin Xie, "An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing", *IEEE Transactions on Information Forensics and Security, June 2016*.

[11]. Shulan Wang, Kaitai Liang, Joseph K. Liu, Jianyong Chen and Jianping Yu, Weixin Xie, "Attribute-Based Data Sharing Scheme Revisited in Cloud Computing", *IEEE Transactions on Information Forensics and Security, August 2016.*

[12]. Vahid Sedighi, Rémi Cogranne and Jessica Fridrich, "Content Adaptive Steganography by Minimizing Statistical Detectability", *IEEE Transactions on Information Forensics and Security –Feb. 2016.*

International Journal of Theoretical & Applied Sciences, Special Issue 10(1a): 19-23(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Middleware Framework for Secure and Trustable Routing using ADDRP in WSNs

***Edwin Rajesh. A*[1] *and Dr. Ponmuthuramalingam. P*[2]**
[1]*Assistant professor, Department of Computer Science,*
CSI Bishop Appasamy College of Arts and Science, Coimbatore-641018 (Tamilnadu), India
[2]*Professor & COE, Govt. Arts College (Autonomous), Coimbatore – 641018 (Tamilnadu), India*

ABSTRACT: Secure knowledge transmission and energy efficiency are one among the foremost outstanding problems for wireless sensor networks (WSNs) are more and more being deployed in security-critical applications. Due to their inherent resource - unnatural characteristics, they are liable to varied security attacks, and a black hole attack may be a form of attack that seriously affects knowledge assortment. Combined knowledge is transmitted during a path exist of connected links. All previous end-to-end routing protocols propose solutions within which every n each link uses a combine wise shared key to guard knowledge. To overcome that challenge, a lively detection-based security and trust routing scheme named ActiveTrust is planned for WSNs. the foremost necessary innovation of ActiveTrust is that it avoids black holes through the active creation of the variety of detection routes to quickly detect and procure nodal trust and so improve the information route security. a lot of significantly, the generation and therefore the distribution of detection routes are given within the ActiveTrust scheme, which may totally use the energy in non-hotspots to make as several detection routes as required to attain the specified security and energy efficiency. Each comprehensive theoretical analysis and experimental results indicate that the performance of the ActiveTrust scheme is best than that of the previous studies. ActiveTrust will considerably improve the information route success chance and skill against black hole attacks and may optimize network lifetime.

Keywords: Wireless sensor networks, Security, Attacks, Trust, black hole attacks, Route, Active Detection data Routing Protocol (ADDRP),optimize network lifetime.

I. INTRODUCTION

Wireless Sensor Networks are dynamic and may contain numerous kinds of device nodes. The surroundings are heterogeneous in terms of each hardware and computer code. Therefore, sensors should have their power provides turned off after they aren't in use high reserve energy. Attributable to this limitation, an important issue becomes a way to prolong the lifetime of WSNs whereas additionally reassuring the service quality of coverage. With radio waves, distances may be short, like many meters for tv or as way as thousands or maybe several kilometres for deep-space radio communications. It encompasses numerous kinds of mounted, mobile, and moveable applications, as well as two-way radios, cellular, personal digital assistants (PDAs), and wireless networking. Alternative samples of applications of radio wireless technology embrace GPS units, garage door openers; Router might offer property inside and between enterprises and the web or between internet service providers (ISP) networks. The most powerful routers are sometimes found in ISPs. To produce a sensible resolution to those issues, our cryptography-based security technique isn't enough. As a result of the attacker will take the encryption/decryption keys, once the actual node is compromised and may interrupt any data suffered the node. Additionally, to traditional security issues like secure routing and secure knowledge aggregation, security mechanisms deployed in WSNs conjointly ought to involve collaborations among the nodes because of the decentralized nature of the networks and absence of any infrastructure. Since the network makes selections looking at the nodes detected knowledge. As a result, the system can totally fail and construct wrong selections. Hence, observe and avoid the attack is nice significance for security in WSNs.

II. RELATED WORDS

In first, technical challenges and style principles are introduced regarding hardware development, system architectures and protocols, and computer code development. Specifically, radio technologies, energy harvest techniques, and cross-layer style for IWSNs are mentioned. additionally, IWSN standards are given for the system owners, who decided to utilize new IWSN technologies for industrial automation applications. Vipul Sharma et al planned a technique for the detection and suppression of black hole attack in Leach based mostly sensor networks. The aim of this analysis work is to advance a mechanism that may observe and overcome the result of black hole attack in a sensor network. The paper proposes a lively detection routing of information for higher security and trust. the most goal of the scheme is to make sure that the nodal

knowledge safely reaches the sink and aren't blocked by the black hole.
The demerit during this paper was it'll not find the sensor nodes as a black hole node. Barleen Shinh projecteda method to observe and isolate the black hole attack. A detection route confirms to a route while not knowledge packets whose goal is to satisfy the individual to launch an attack that the system will realize the attack behavior stick the black hole location. Jian-Ming Chang et al. developed a brand new mechanism known as Cooperative bait detection scheme (CBDS) for detection malicious nodes in MANETs below gray/collaborative black hole attacks. During this approach, the supply node stochastically selects the associate adjacent node with that to get together, that the address of the adjacent node is employed as bait destination address to bait malicious nodes to send a reply message.

III. WIRELESS SENSOR NETWORK

The WSN is made of "nodes" – from many to many a whole bunch or perhaps thousands, wherever every node is connected to at least one (or generally several) sensors. the value of sensor nodes is equally variable, starting from any to many dollars, looking on the complexness of the individual sensor nodes. Size and value of the constraints in this sensor elements are a sensor node are lead to the corresponding constraints in the resources like as energy, memory, processor speed and the communications information measure. The topology of the WSNs will vary from an easy star network to a complicated multi-hop wireless mesh network. that the cross-layer may be accustomed build the optimum modulation to enhance the transmission performance, like rate, energy efficiency, QoS (Quality of Service), etc.

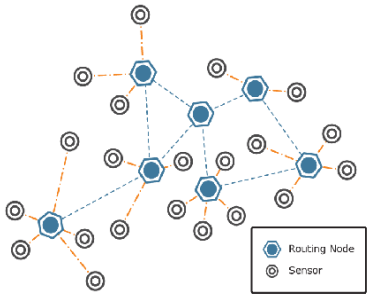


Fig. 1. Sensor Network.

Sensor nodes may be imaginary as little computers that are extraordinarily basic in terms of their interfaces and their elements. They act as an entryway between sensor nodes and therefore the user as they usually forward knowledge from the WSN on to a server. different special elements in the routing primarily based networks arrouters, designed to compute, calculate and distribute the routing tables.
Sensor node: Although wireless sensor element nodes have existed for decades and used for applications as various different types of earthquake measurements to warfare, the trendy development of little sensor nodes dates back to the 1998 good dust project and therefore the NASA sensor Webs Project one among the objectives of the good dust project was to make autonomous sensing and communication among a metric capacity unit of space.
Sensor: Sensors live physical knowledge of the parameter to be monitored. The continual analog signal created by the sensors is digitized by an analog-digital converter and sent to controllers for the additional process. A sensing element node should be little size, consume extraordinarily low energy, operate in high volume trical densities, be autonomous and operate unattended, and be adaptive to the surroundings.
Routing Node: Routing is that the method of choosing best methods in an exceedingly network. in the past, the term routing additionally meant forwarding network traffic among networks. However, that latter perform is best represented as forwarding. Routing is performed for several types of networks, as well as the telephone network (circuit switching), electronic knowledge networks (such because of the Internet), and transportation networks. this text thinks about primarily with routing in electronic knowledge networks exploitation packet switch technology.

IV. EXISTING SYSTEM

Single-path routing may be an easy routing protocol, however, is well blocked by the attacker. Therefore, the foremost natural approach is via multi-path routing to the sink. although there's an attack in some route, the information will still safely reach the sink. Multi-path routing protocols may be classified into two categories looking on whether or not the information packet is split. One is multi-path routing while not share division. The other is multi-path routing with share division, i.e., the packet is split into shares, and totally different| completely different shares reach the destination via different routes. Non-share-based multi-path routing. There are totally different multi-path route construction strategies. Another paper proposes a multi data flow topologies (MDT) approach to resisting the selective forwarding attack. within the MDT approach, the network is split into two data flow topologies.
The basic plan of the unfold algorithm is to remodel a secret message into multiple shares, that is named a (T, M) threshold secret sharing scheme. The M shares are delivered by multiple freelance methods to the sink specified, although a little range of shares is born, the key message as an entire will still be recovered.

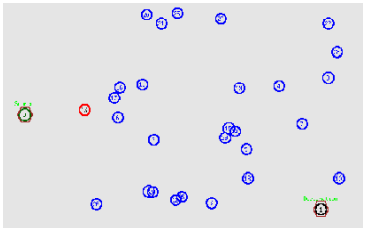


Fig. 2. Drawbacks in the Existing.

The advantage of this algorithm is that through multi-path routing, every path routes just one share, and therefore the attacker should capture a minimum of T shares to revive nodal data, that will increase the attack issue.

Thus, the privacy and security may be improved. within the higher than analysis, the multi-path routing algorithms are deterministic specified the set of route methods is predefined below a similar topology. This weakness opens the door for numerous attacks if the routing algorithm is obtained by the individual. For the weakness mentioned higher than, proposed four random propagation strategies: random propagation (PRP), directed random propagation (DRP), non-repetitive random propagation (NRRP), and multicast tree assisted random propagation (MTRP).

In multi-to-one knowledge assortment WSNs, we tend to argue that for traditional "slicing and assembling" or multi-path routing 0techniques, sliced shares can merge within the same path with high chance, and this path may be simply attacked by black holes. so a Security- and Energy-efficient Disjoint Route (SEDR) scheme is planned to route sliced shares to the sink with randomised disjoint multipath routes by utilizing the obtainable surplus energy of sensor nodes. additionally proposes a resilient trust model, Sensor Trust, for hierarchical WSNs. Introduces the conception of attribute similarity to find probably friendly nodes among strangers.

V. PROPOSED SYSTEM

Proposed an Active Detection data Routing Protocol (ADDRP) during this method. The Active Detection protocol algorithm is employed to seek out the neigh bore node of during this network. Calculation of Nodal Trust algorithm - throughout knowledge routing and detection routing, each node can perform a nodal trust calculation to help in black hole rejection.

Upon detection of a happening, a sensor node can generate messages, and people messages should be sent to the sink node. we think about that link-level security has been established through a typical cryptography-based protocol. Thus, we think about a link key to be safe unless the individual physically compromises either aspect of the link.

The adversaries model: we think about that black holes are fashioned by the compromised nodes and can unselectively discard all packets gone to stop information from being sent to the sink. The somebody has the flexibility to compromise a number of the nodes. However, we think about the individual to be unable to compromise the sink and its neighboring nodes.

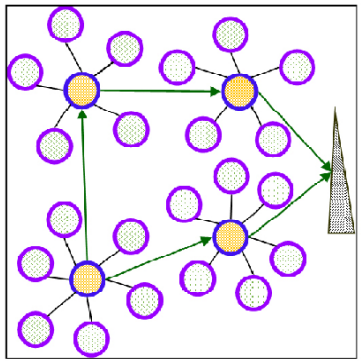


Fig. 3. Example of proposed system.

The data assortment has higher security performance and powerful capability against black hole attacks. the most goal of our scheme is to confirm that the nodal knowledge safely reaches the sink and aren't blocked by the black hole. Thus, the scheme style goal is to maximise the ratio of packets with success reaching the sink.

Thus, the system will lower the trust of suspicious nodes and increment the trust of nodes in successive the routing routes. Through active detection routing, nodal trust may be quickly obtained, and it will effectively guide the information route in selecting nodes with high trust to avoid black holes. The active detection routing protocol is that the scheme, the supply node at random selects an undetected neighbor node to form a lively detection route.

The routing protocol is comparable to common routing protocols in WSNs; the distinction is that the route can choose a node with high trust for the following hop to avoid black holes and therefore improve the success ratio of reaching the sink. If there's not a node among all neighbours nearer the sink that has trust higher than the default threshold, I'll report back to the higher node that there's no path from a to the sink.

The higher node, operating within the same manner, can re-select a special node from among its neighbors nearer the sink till the information are routed to the sink or there's once and for all no path to the sink.

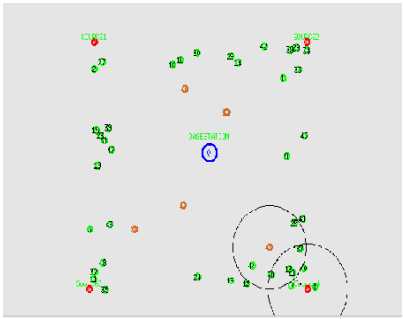


Fig. 4. Secure routing.

In the ActiveTrust scheme, the trust calculation ought to meet the subsequent condition. If the node is found to be malicious within the latest detection, then its trust ought to be below the edge, and also the node won't be chosen for later routing. The trust calculation supported the remaining energy node. The signature verification and unidirectional hash chain offer secure communication within the network. If the malicious node returns to the traditional node, it desires many detections to require it into routing consideration; The core plan of knowledge routine is that once any node receives a knowledge packet, it selects one node from the set of candidates nearer the sink whose trust is larger than the planned threshold because the next hop. we've got planned a completely unique security and trust routing scheme supported active detection, and it's the subsequent wonderful properties High successful routing chance, security and measurability. during this method as a lot of detection, routes are performed, a quantity does black nodes detected grows quickly; once the quantity of deployed black nodes.

- High Throughput.
- High security.
- No packet dropping.
- Network lifetime will be high once compare with the existing system.
- High Pack delivery radio.

VI. SYSTEM ARCHITECTURE

The current trust-based route ways face in getting trust. Energy efficiency. because energy is incredibly restricted in WSNs, in most analysis, the trust acquisition and diffusion have high energy consumption, that seriously affects the network lifetime.

Because it's troublesome to find malicious nodes, the protection route remains a difficult issue. The activetrust route protocol has higher energy efficiency. Energy is incredibly precious in WSNs, and there'll be additional energy consumption if active detection is processed.

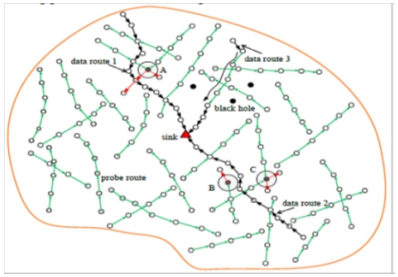


Fig. 5. ActiveTrust Scheme.

Therefore, in a previous analysis, it absolutely was not possible to imagine adopting such high-energy-consumption active detection routes.

The attacker's behavior and site, similarly as nodal trust, may be obtained and accustomed avoid black holes once process real knowledge routes. To the simplest of our information, this can be the primary planned active detection mechanism in WSNs.

The most important difference between activetrust and former analysis is that we produce multiple detection routes in regions with residue energy; as a result of the attacker isn't aware of detection routes, it'll attack these routes and, in therefore doing, be exposed. during this approach, the attacker's behavior and site, additionally as nodal trust, may be obtained and wont to avoid black holes once process real knowledge routes.

To the most effective of our data, this can be the primary projected active detection mechanism in WSNs. The activetrust route protocol has higher energy efficiency. Energy is incredibly precious in WSNs, and there'll be a lot of energy consumption if active detection is processed.

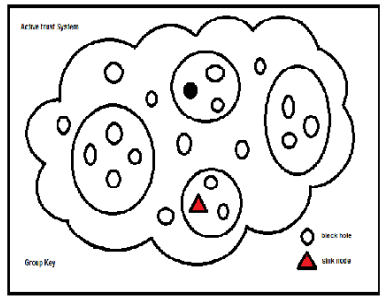


Fig. 6. ActiveTrust Architecture.

Therefore, the activetrust scheme takes full advantage of the residue energy to form detection routes and

makes an attempt to decrease energy consumption in hotspots (to improve network lifetime). Those detection routes will detect the nodal trust while not decreasing lifetime and therefore improve the network security.

VII. CONCLUSION

In this, ActiveTrust secure and trustable routing in wireless sensor network is with success reached method. A secure and trustable hierarchical routing for big scale WSNs has been planned to avoid black holes with efficiency and to transmit knowledge firmly. The activetrust scheme absolutely uses residue energy to construct multiple detection routes. They have given trustable routing and also the security they have given action. it'll additionally improve each the energy efficiency and also the network security performance. ActiveTrust will considerably improve the information route success chance and skill against part attacks and may optimize network lifetime. during this WSNs improve the high throughput, high security, Network lifetime are going to be high once compare the present system and high delivery existing radio.

REFERENCES

[1]. Liu, A., M. Dong, K. Ota and J. Long, 2015. PACK: AN efficient scheme for selective forwarding attack detection in WSNs. *Sensors Journal*, **15**(12): 30942-30963.

[2]. Z. Zheng, A. Liu, L. Cai, et al."Energy and Memory efficient Clone Detection in Wireless sensing element networks," IEEE Transactions on Mobile Computing.vol. **15**, no. 5, pp.1130-1143,2016.

[3]. Lai, C., H. Li, R. Lu and X.S. Shen, 2013. SE-AKA: A secure and efficient cluster authentication and key agreement protocol for LTE networks. *laptop Networks*, **57**(17): 3492-3510.

[4]. P. Zhou, S. Jiang, A. Irissappane, et al. "Toward Energy-Efficient Trust System Through Watchdog optimisation for WSNs," *IEEE Transactions on info Forensics and Security*, vol. **10**, no. 3, pp. 613-625, 2015.

[5]. J. Chen, X. Cao, P. Cheng, Y. Xiao, and Y. Sun, "Distributed collaborative management for industrial automation with wireless [3] sensor and mechanism networks," *IEEE Trans. Ind. Electron.*, Vol. **57**, No. 12, pp. 4219–4230, Dec. 2010.

[6]. X. Cao, J. Chen, Y. Xiao, and Y. Sun, "Building-environment management with wireless sensor and mechanism networks: Centralized [4] versus distributed," IEEE Trans. Ind. Electron., Vol. **57**, No. 11, pp. 3596–3604, Nov. 2010.

[7]. R. E. Mezouary, A. Houmz, J. Jalil and M. E. Koutbi, "PRoPHET-RAIP5: a brand new approach to secure routing in wireless sensor networks," 2015 International Conference on Wireless Networks and Mobile Communications (WINCOM), Marrakech, 2015, pp. 1-6.

[8]. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in the good grid," IEEE Trans. Ind. Electron., vol. **57**, no. 10, pp. 3557-3564, Oct. 2010.

[9]. And I. Hubaux P. J. and. Knightly W. E, 2008."Impact of Denial-of-Service Attacks on Ad-Hoc Networks," IEEE-ACM Transactions on Networking, vol. **16**, no. 4, pp. 791-802, He Q. , Wu D., Sori P. K, 2004.

[10]. Akyildiz, IF, Su, W, Sankarasubramaniam, Y & Cayirci, E 2002, 'A survey on sensing element networks', *Communications Magazine, IEEE*, vol. **40**, pp.102-114.

[11]. Michael, L, Raymer, William, F, Punch, Erik, D, Goodman, Leslie, A, Kuhn, & Anil, K, Jain 2000, 'Dimensionality Reduction exploitation Genetic Algorithms,' *IEEE Trans. evolutionary Computation*, vol. **4**, no.2, pp.164-171.

[12]. Yuxin Liu, Mianxiong Dong Ota, Kaoru and Anfeng Liu," ActiveTrust in the Secure and Trustable Routing in Wireless Sensor Networks", *IEEE transaction on data forensics and security*, Sep 2016, Vol. **11**, No.9.

[13]. Jamal N. AI-Karaki, Ahmed E. Kamal, "Routing techniques in wireless sensor networking: A Survey", proceedings by the I-CUBE initiative of Iowa state university, IA 50011,2004.

[14]. Vipul Sharma, Kirti Patel, Ashish Tiwari "Detection and Suppression of Blackhole Attack in Leach primarily based sensor network", *International Journal of technology and Applications*, Vol **5** (6),1873-1877, 2014.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 24-29(2018)

International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Social Media Marketing on Travel and Tourism Industry

Dr. S. Preetha[1] and V. Bharathi[2]

[1]Assistant Professor, Department of Computer Science,
Sri Ramakrishna College of Arts & Science for Women, Coimbatore (Tamilnadu), India
[2]Research Scholar, Department of Computer Science,
Sri Ramakrishna College of Arts & Science for Women, Coimbatore (Tamilnadu), India

ABSTRACT: Travel and Tourism is one of the industries that is mostly influenced by digital development. Tourism and accommodation companies are among the first ones to make use of digital marketing techniques in their practices to engage their customers to provide the feel at home experiences. Digital Marketing plays a vital role in every business in the Tourism industry. As tourists and travelers are looking for instant information, which is provided from online business websites, mobile apps and social media. There is drastic change in customer's behavior in association with the technologies of the internet. A growing number of travel agencies have changed from their traditional system to online travel agencies. Travel &tourism service providers build their business by offering great travel deals online by digital medium. This paper focus the impact of the digital marketing on travel and tourism industry.

Keywords: Digital Marketing, Internet Marketing, IT for Tourism, Online Bookings, Social Media Marketing, Travel & Tourism.

I. INTRODUCTION

In this internet era, the rules of all the business have been redefined all over the world. The impact of digital transformation in every business area has been grown tremendously. And there is no exception for travel and tourism industry also. This digital transformation has become the key focus area of business growth now-a-days. Internet has stepped in as a mode of communication and has evolved into mandatory to maintain and develop the business into next level for long term relationship so that it provides opportunity for the brands to interact with their customers directly. Travelers wander quickly from one place to other place. They browse the internet for wider range of information of hotels and travel agencies on multiple devices so as to get personalized service and easy booking experience.

The travel companies incorporated digital marketing strategies to reach their right audience for improving the better exposure and gain insights of their brand. The ultimate goal of digital marketing to reach the target audience and improve the brand or business to lead more success. The customer should be able to interact with the business.

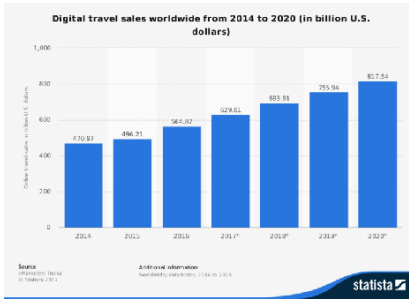
II. UNDERSTANDING DIGITAL MARKETING

Digital marketing is the marketing of brands or products in the internet (electronic media) which uses different methods and channels to reach their audience. The digital marketing can be done in various mediums to get popularize and to get conversions for a good startup of a business. The various forms include

1. Search Engine Optimization(SEO)
2. Search Engine Marketing(SEM)
3. Pay Per Click Advertisement(PPC)
4. Content Marketing
5. Social Media Marketing
6. Influencer marketing and many more.

The need of the customer is understood whenever digital campaign is launched which helps to track the user behavior in life time, engagement and conversion of the customers. With the evolution of the digital marketing in the travel and tourism industry, the new ways to reach the target customers by blogging, paid search campaigns, content marketing, search engine optimizations, social media like twitter, Facebook, Instagram. These practices shapes the travel and tourism industry to get more digitalized.

Traditional marketing helps to understand the audiences from the analysis of data collected in market research. Transformation of traditional marketing with internet technology gathers detailed information on customer's behavior and their characteristics in electronic media. This business transformation requires careful planning, adaption and inclusion of many digital trends and strategies which challenge the business for better growth with the consistent flow of data integration.



In the past few years, different industries like finance, retail, healthcare, media and entertainment transformed digitally, whereas the extreme digital transformation happened in travel and tourism industries. The travel and tourism industry is competitive field with the driving force of digital era, it's important to the business to embrace the changes to get along with the customer's journey for strong adaption of its brand.

Some of the digital marketing techniques for travel and tourism industries to consider:

- Rich content
- Conversational Marketing
- Chatbots
- Artificial Intelligence
- Advertisements
- Live Streaming Videos
- Customer Reviews
- Email Marketing
- Social Media Marketing
- Virtual Reality
- Augmented Reality

Rich Content: The main objective is to create gripping content to achieve the goals and promote it effectively.

In content marketing strategy, rather producing expansive quantities of normal content, we focus on high quality content that uplifts the brand to great effect. The content marketing strategies include:

- Promoting the high quality on social platforms
- Gathering sharable and relatable information.
- Focusing on SEO for the high quality content
- Content have to be used for further reference and for the research purposes.

Conversational Marketing: Live Chats and messaging Apps are widely used and becoming popular. From this interaction, the customer can clarify the queries regarding the travel and sometimes with the friendly conversations, it might end in the booking of a Holiday trip also.

Chatbot: Chatbots includes both customer and customer service agents which provides excellent and efficient way of customer service to the user engagement. The chatbots have become new trend now because they make their own way in digital marketing via messaging applications and social media.

Advertisements: Whenever the website or any app is opened, various advertisements are bombarded in our desktop or mobile. So the message which has to be conveyed to the customer should be effective. The advertisements have to be designed according to the devices either desktop or mobile.

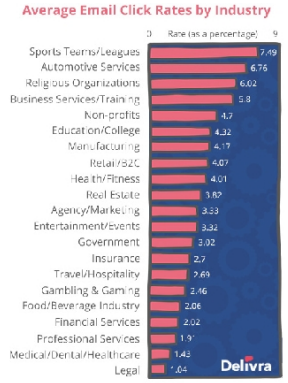
Artificial Intelligence: Another important key word in current digital marketing trends in tourism is Artificial Intelligence. In order to perform the traditionally human task by the computers, Artificial intelligence is used as they require intelligent actions. Today, restaurants and hotels can provide all the information required with AI powered apps and technologies. Few hotels started using artificially intelligent robots which provides most of the tourist information and the directions in response to the human's speech.

Videos: Engaging the customers with the effective content about the travel and tourism industry in the shorter span. The text based content will not have much effect when compared to videos. Exploration for the promotion of new bands provides the excellent results on their conversion. Social Media like Facebook, Snapchat and Twitter has made the Videos to the bigger level. Live Video is one of the popular marketing tool for the brand or business where they work with influencers and take viewers to live through an event. Travel social media people use this tool to show their

visitors the hotel place or destination they are staying in which obviously provides the direct view of the stay.

Customer Reviews: Before planning for a travel, any customer would like to read the reviews to book for a trip. The online reviews reflects the customer experience in the stay of a hotel or travel or boarding in a flight and many more. Psychologically, when we hear a good feedback about a restaurant or a hotel or any place, we are excited to get into the place where we have a social proof of that particular place is desirable.

E-mail Marketing: E-Mail marketing establishes the opportunity to provide potential customer at the appropriate time with minimum cost. These activities are measurable that offers a way for future marketing strategies. In travel and tourism industry, email marketing helps in such a way which enables the users to get to know the latest offers provided by the travels, hotels, membership clubs, and some special deals for the holiday packages.



Building the special email marketing strategies include

- Creation of strategy
- Understanding the importance of email
- Establishing subscriber personas
- Identifying the life time value of a guest
- Employing RFM (Recency, frequency, monitory) analysis.

Virtual Reality: There are so many to showcase in travel and tourism industry like venue of the destination, amenities in the hotel and restaurants, location for open space to accommodate for a wedding or conference and many more. These destinations can be seen virtually through a technology called virtual reality.

This technology helps the customers to feel as if they present physically in the destinations in a digitally created environment including different sights and sound.

Even the hotel booking websites are leveraging the potential of allowing the customers to use VR technology to showcase the virtual recreations of the hotels. Rather reading pamphlet or description of a travel, customers can feel the experience before they book them.

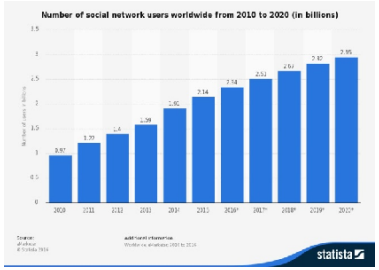
Augment Reality: Augmented reality is going to be the trending fizzy word in 2018. Augmented Reality is an innovative way of connecting customers and increasing their engagement. The augmented reality is close to the concept of Virtual reality. AR uses digital technology to alter the experience of real-life surroundings. They are enhanced is such a way when viewed through an AR compatible device. Brands can use augmented reality in social media also where the virtual experience are available endlessly.

III. SOCIAL MEDIA MARKETING

Understanding Social Media: Social Media is a platform where the people can able to freely interact with one another, or to any organizations and companies. In other words, it provides a platform that allows the user to share their content in form of reviews, photos, ratings, stories and many more. Internet users can typically access the social media through the web based technologies. **Social media network is a gateway for the organizations and individuals to work together, learn, even buy and sell Buying and selling gains a huge importance in social media marketing.

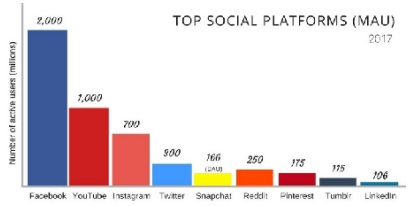
Social media are the one which is based on computer mediated technologies that makes the easier way for better communication between the individuals and the organization. Social media is the collective of online communication facilitates to content sharing, interaction, collaboration and community-based input.

The way in which the companies communicate to the target audience have changed extremely is because of the social media. The prominent rise of internet and various social channels has established a drastic change in marketing trends. Undoubtedly, digital marketing has created the revolution in the travel and tourism industry. The new ways that customers use social media to make travel decision have effected travel and tourism marketing from the beginning till the end. The travel and tourism industry have already replaced the traditional communication and transform to digital marketing. The travel industries can make use of social media marketing tools to promote their brand at low cost and tend to attract more visitors to the page.



People often write reviews, recommendations, post photos will motivate the other travelers to make further decisions and provides a wider platform to showcase their travel experience of the other guests also. Social media is definitely beneficiary to the travel and tourism industry. When correct form of digital tourism marketing is practiced which can lead the business to attain more profit in a higher scale. The most appropriate social media platform for a travel and tourism industries are:

- Facebook
- YouTube
- Twitter
- LinkedIn
- Instagram
- Pinterest
- Reddit
- Snapchat
- Tumblr



Facebook has emerged as household name and with the below report, it is noted as Facebook is the highly effective platform for the promotion of the brand, Even influencer marketing strategies can be incorporated under social marketing to influence more number of users and to encourage them to travel and explore new places.

From the above chart representation, LinkedIn has started making improvements currently, offering new platform for B2B and B2C clients.

From Track Maven, which is one of the marketing software companies, have analyzed and revealed that a better engagement on can be noticed on Instagram than on any other social media. Companies can make use of the available digital tourism marketing trends for the promotion on the brand in many ways. Instagram can be used for live videos and stories. Snapchat for snap stories. For live streaming, Facebook can be utilized in the effective manner.

The online research have found that travel companies which interact their target audience on social media can able to increase their brand image more effectively. When there are more engagements, automatically number of followers, reviewers will get increased which results in higher return of investment. Round the clock, travel and tourism industries should present on their social platforms of Facebook, Twitter or whatever the platform.

Social marketing can bring more conversions that any other expensive digital marketing as even a common man or business person does a research on the travel before the booking the deal. Therefore the services have to be kept up to the mark and ready to take require action whenever the negative reviews falls on the business otherwise the entire marketing strategy will go down.

Here are the few ways that travel and tourism industry have been impacted:

1. Research before the travel: Travelers go internet to research their destination and hotel stays for their future holidays. Research shows that social media emphasize on travel plans. Some of the factors which influence the research prior to their travel are

Online Customer Reviews: Before planning for a travel, any customer would like to read the reviews to book for a trip. The online reviews reflects the customer experience in the stay of a hotel or travel or boarding in a flight and many more. Psychologically, when we hear a good feedback about a restaurant or a hotel or any place, we are excited to get into the place where we have a social proof of that particular place is desirable.

Online Reviews is one of excellent ways to identify any problems in the services which they provide and allow them to take necessary measures to handle these situations, finally which would lead to more customer satisfaction. These reviews have be monitored and responded without any delay to avoid further problems. This provides a way to engage the customers to talk to them irrespective of positive or negative comment.

When the review is positive, the company is responsible to respond with the "Thank You" note to the customer back. At the same time when the review is negative, it is very important to inform the customers how to fix the issues and addressing the areas of the concern. We must make sure a plan or concerned

person to deal with these online reviews alone for better service.

Trip Advisor is one of the websites leading top in the review of the customer's experience which helps the tourism business to get their reviews, increased rating levels and rankings. The tourism ventures can have tie up with the Trip advisor to have the get the reviews collected and sending them via email to the firms, also sometimes they promote the positive reviews on the social media to increase the ROI.

Many people started looking for reviews than ever before, and from the reviews they have found that reviews depends on following factors:

- Local business
- Trust and influence
- Positive Reviews
- Fake Reviews

Here are the few tips in managing the online customer reviews:

Using Online Reputation Management Software:

Software like Grade. Us and Bright local facilitates the customer to write reviews. These types of software enables the customers to review the sites that the business profile provides. This also guides the unhappy customer and unsatisfied customer to a service recovery.

Ask every customer for feedback: Whether customer's feedback is good or bad, the review has to be asked to every customer, so that all the flaws will come out publicly, provides a free insight on how to improve a business.

Responding to reviews: If the customer's feedback is positive, respond the review with honest note as "Thank You!" If the customer's feedback is negative, it is better not to react on it immediately, and respond with the necessary actions with an apology note in humble approach. This method of solving problems engages with the customer and will result in more positive feedback.

Social Sharing: People who are travelling will always like to post their photos and videos taken in their travels. In fact, 52% of the Facebook report confirms that friend's photos and videos have inspired the future travel plans. Most of the travelers update their Facebook status while on vacation. This means that it is very significant to travel and tourism brands to establish relationships with the customers on social media to motivate with the positive approach. Usage of hash tag has a tremendous response, which significantly improves the business.

Over 97% of the travelers share their photos and videos online, that facilitates to inspire the potential guests. Many resorts and hotels have started to run their campaigns and social contests to ensure some sort of credit from their customer's social activities.

2. Enhanced Customer Service: Social Customer service is the practice of leveraging consumer support through social media channels such as Facebook and Twitter. The customer services helps to track the brand to become aware, and when necessary to provide help to confused or unsatisfied customers. Apart from this, customer service representative can be on the part of the travel firm for the company's social media interactions. The customer service helps to maintain and develop true relationships with the customers when compared to email or phone. Also, Social media is a proactive way for the travel industries to optimize the customer retention. Listening through social media can definitely create an experience which will delight the customers.

3. Reshaping travel agencies: Social media has played a vital role in travel agency model. The self-service booking and availability of the information has pushed the travel agencies to transform into digitalized. The online travel agencies (OTA) have a tremendous growth than any other industries. Booking for buses, trains and flights to hotels, resorts and restaurants dominate the travel industry. The travel agency should create OTA strategies so that they have personalized touch with the potential guests.

4. Changing Loyalty programs: Loyalty programs have a great impact on the travel industries and social media emphasize on the loyalty programs business model. Acquiring new customers is very expensive than retaining existing customers. Many customers understand that they get tremendous response when they share their opinions in the social media. In order to exchange for the loyalty points, more than 25% of the travelers take part in the loyalty programs.

It is very much easier for the hotels to discover the potential guests and reward them with the availability of the current technologies especially with the hash tag tracking process. When the loyal guests share the benefits offered by business loyalty programs on social media, other customers could feel that the benefits are easily attainable which encourages them to participate. Most travelers decide their travel plans online reviews and social media shares. The travel and tourism brands builds a positive brand reputation by increasing the brand loyalty, and inform about the deals and promotions of the travel business.

Social media networks collaborated with the new technologies keeps track of the loyalty program for the customers and informs about the deals and promotions on that particular time itself. Specifically in Facebook, any deals and promotions are informed to the customers when the brand's page is liked or the brand's videos and photos are shared to their own community. This makes other travelers to get into the site and expect to book a holiday.

These types of promotions will definitely increase the Return of the investment. Even the Virtual Reality and Artificial Intelligence can be used in the Facebook pages to increase when the promotions are informed.

IV. TOOLS FOR SOCIAL MEDIA LISTENING AND MONITORING

In this world of digital channels, the revenue generated in the social media platforms continue to strike new record. Facebook reaches over 2 billion users whereas the Instagram is expected to reach 1 billion users. Social media users leverage the following tools to reach the target customers either by organic or by paid means. There are few main tools for social media,

Social Drift – One of the significant tool for the Instagram followers

Buffer - To manage multiple social media accounts from one place which helps to analyze the channel performance, schedule social media posts and monitor conversations.

Owlmetrics – helps to get the information about the growth and engagement of the followers through the dashboard.

Sproutsocial – It is wise chance for the social media solution in enterprise level

Adespresso – Integrating Facebook with the marketing automation platform which provides up-to-date list of contacts from custom audiences.

Sprinklr – Social media cloud designed to help enterprise organizations for business analytics.

Adobe spark post – Facilitates high quality social media images.

Funnel.io – Gathers advertising information in one simple dashboard.

Salesforce Social Studio –Powerful social media management solution for enterprise businesses.

Splice – User friendly tool to edit the social media video content.

V. CONCLUSION

Now-a-days, the young people make use of social media extensively, and this scenario have both positive and negative impact on every business especially in travel and tourism sector.

Even slightest negative feedback on high quality enterprise will impact more. So that high quality enterprises will hire a dedicated staff to manage its accounts separately. The brand has to be active on the social media page to establish an engaging post which should inspire other followers to travel. Usage of hast tag will increase the visibility of the travel brand. When the travelers post the photos and videos in their pages, the brand can repost the same in their respective social media page with their authentication in order to inspire other travelers.

Social media is an effective tool for high quality enterprises and high potential customers to provide high quality service. The social media and tourism industry is undoubtedly perfect combination to insight greater heights in the travel business.

REFERENCES

[1]. 10 Social Media Tools You Need in 2018 | Inc.com
[1]. 10 Social-Media Trends to Prepare for in 2018
[1]. 5 Digital Marketing Trends transforming the Travel industry
[1]. 8 Most Important Travel Trends for 2018
[1]. Column: 5 top travel and tourism trends for 2018
[1]. Important Social Media Marketing Trends for the Travel Industry - Digital Hotelier
[1]. How to Leverage Social Media In 2018: A Video Marketing Guide for Brands | Inc.com
[1]. Journal of Travel & Tourism Marketing: Volume 35, No 1
[1]. Local Consumer Review Survey 2017 | The Impact Of Online Reviews
[1]. Mobile marketing statistics 2017
[1]. Travel trends: 2018 looks strong for the tourism industry
[1]. Travel Industry Digital and Media Trends - USDM Digital
[1]. Tourism industry optimistic for growth in 2018
[1]. Tourism marketing trends to tap into in 2018
[1]. Top 10 Travel Marketing Conferences of 2018 – The Pixlee Blog

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 30-33(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Comparative Study of Various Energy Efficient Techniques used in Wireless Sensor Networks

N. Deepa[1] and Dr. D. Devi Aruna[2]
[1]Ph.D. Research Scholar, Department of Computer Science,
Dr. NGP College of Arts and Science, Coimbatore (Tamilnadu), India
[2]Assistant Professor, Department of Computer Science,
Dr. NGP College of Arts and Science, Coimbatore (Tamilnadu), India

ABSTRACT: The wireless sensor networks (WSN) is useful in many real time applications such as medicine, agriculture, military, environmental monitoring, tracking purposes and etc. The security issues and the energy efficiency is a major issue that affects by using WSN. There are number of techniques, protocols are being developed for energy consumption and the security approaches to prolong the lifetime of the network. A maximum utilization of resources need to concern while we can maximize the lifetime of a network routing nodes. In this paper we attempt to comparatively analyze the various protocols and different techniques which are already used in WSN.

Keywords: Energy efficiency, Consumption, security, Routing.

I. INTRODUCTION

Wireless Sensor network is an important area of research while working or discussing about the different communication medium today. Wireless sensors are powered by battery and since most of wireless sensors are distributed in hostile environments. Deficient of energy will enforce the nodes to getting die and useless and finally this phenomenon causes failure of the whole network goals. If all nodes of the network start to send and receivedata directly, this strategy causes a rapid depletion of energy [1, 2]. Due to the wireless dynamics, asynchronized low-power approaches, which do not require time synchronization, have been shown to be more appropriate in real deployments [3]. Sensor networks can be seen as an extreme form of ad-hoc networking with very low power devices [4]. The requirement for energy efficiency suggests that in most cases computation is favoured over communication, as communication is three orders of Magnitude more expensive than Computation [5]. This paper describes the existing routing techniques in WSN and gives an overview about energy efficient routing protocols like ECHERP, PRRP, CELRP, FACOR, LEACH, PEGASIS, HEED, DECA, EEMCRP and HRP.

II. EQUALIZED CLUSTER HEAD ELECTION ROUTING PROTOCOL: ECHERP

This protocol which pursues energy conservation through balanced clustering. Equalized Cluster Head Election Routing Protocol pursues energy conservation through balanced clustering For Energy Efficiency. Used to reduce energy consumption by localizing communication within the cluster and aggregate the data to reducetransmission to the BS. The sensing electronics measure the ambient conditions related to the atmosphere surrounding the sensors and convert them into an electrical signal. ECHERP calculates the combinations of nodes that can be chosen as cluster heads in order to extend the network lifetime. The objective of this project is to simulate and analyses performance of routing protocol by using various metrics like energy usage in terms of network load, throughput, energy efficiency and delay and end-to-end performance characteristics of algorithms. It also minimizes the energy and latency for cluster formation, in order to minimize overhead to the protocol. ECHERP [6] (Equalized Cluster Head Election Routing Protocol), with the purpose of growth of the network lifespan elects a node as cluster head that reduces the total energy intake in the cluster and not the node with the higher energy. ECHERP also accepts multi hops routing system to transfer data to base station. In ECHERP, the BS is predictable to have unlimited energy residues and communication power. ECHERP models the network as a linear system and, using the Gaussian elimination algorithm, calculates the combinations of nodes that can be chosen as cluster heads in order to extend the network lifetime.

III. POSITION RESPONSIVE ROUTING PROTOCOL: PRRP

Position Responsive Routing Protocol (PRRP) is used in Global Positioning Systems (GPS) approach for enhancing energy efficiency of wireless sensor network, This Research results shows a 23% to 25% significant improvement in overall efficiency of WSN. Routing protocol is the main energy expensive operation of sensor networks which utilize maximum energy resources of sensor network.

IV. CLUSTER BASED ENERGY EFFICIENT LOCATION ROUTING PROTOCOL: CELRP

The main aim ofthis routing protocol in wireless sensor network is to find a way to improve energy efficiency for reliable transmission of sent data to the base station. Most of the routing protocols can be classified according to the hierarchical structure as and location based [7]. The two kinds of hierarchical routing tactics distributed routing and centralized routing..It shows good performance for average energy per packet compared to PRRP. However, after a short span of time PRRP shows a noteworthy improvement in it with the increased number of data periods. The performance of the CELRP was network life span, throughput. To prove the efficiency of energy in CELRP which was compared toBase Station Controlled Dynamic Clustering Protocol (BCDCP) and Greedy Perimeter Stateless Routing (GPRS). Since it was used in a small scale network and not so much in a larger scale network because it uses a lot of energy for long distance wireless communication.

V. FUZZY ANT COLONY OPTIMIZATION ROUTING (FACOR)

This paper proposes a novel approach called fuzzy ant colony based routing protocol (FACO) using fuzzy logic and swarm intelligence to select optimal path by considering optimization of multiple objectives while retaining the advantages of swarm based intelligence algorithm. Simulation results show that the proposed protocol is superior over existing swarm intelligence based routing protocols for routing in MA.

VI. ANT COLONY OPTIMIZATION

The routing problem is a very important part in this kind of networks that need to be considered in order to maximize the network life time.This protocol was studied by simulation for several Wireless Sensor Network scenarios and the results proved that it reduced communication load and increased the energy savings. This algorithm addresses the adaptation of the collective behaviors observed in natural ant colonies for routing in wireless sensor network WSNs, ant swarms usually collectively achieve adaptive, scalable, and robust optimized paths between the net and the source of food with little intelligence and capacities at each individual which is very suitable from WSNs perspective which are composed of small sensors with limited capacities and resources in hostile and unpredictable environment.

VII. LOW ENERGY ADAPTIVE CLUSTERING HIERARCHY: LEACH

LEACH is a protocol that tends to reduce energy consumption in a WSN. It is a cluster-based, energy-aware routing protocol specifically designed for sensor network. The LEACH includes distributed clusterinformation, local processing to reduce global communication andrandomizedrotation of the cluster heads. Energy efficiency is very high in comparison to other routing protocol. This type of protocol is implemented in hierarchicalclustering. This improves the energy conservation, throughput, and network lifetime. It helps to prevent energy drain for the same sensor nodes which has been elected as the cluster leader, using randomization for each time cluster head would be changed. The cluster head is responsible for collecting data from its cluster members and fuse it. Finally each cluster head will be forwarding the fused data to the base station. When compared with its previous protocols LEACHhave shown considerable improvement.

VIII. POWER-EFFICIENT GATHERING IN SENSOR INFORMATION SYSTEMS: PEGASIS

Power-Efficient Gathering in Sensor Information Systems (PEGASIS) is an energy efficient protocol [8], which provides improvements over LEACH Its is an optimized chain based protocol. The PEGASIS protocol achieves between 100 to 300 % improvement when 1%, 20%, 50% and 100% of nodes node die compared to the LEACH protocol. The drawback of this protocol is when head node is selected, the energy level is not considered. PEGASIS energy efficient is achieved by transmitting the data to only one of its neighbor node. The Shortest path to the super cluster further reduces the power consumption. Compression techniques for reduced data fusion cost.

IX. HYBRID ENERGY EFFICIENT DISTRIBUTED PROTOCOLS: HEED

This protocol is designed for heterogeneous wireless sensor network and moreover used in clustering routing protocol. This protocol guarantees that every sensor is part of just one cluster, and the cluster heads are well-distributed. So it enhanced the network life-time by distributing energy consumption. It also minimizing the control overheadand it does not require any special node capabilities, assumptions about node distribution and also Operates correctly even when nodes are not matched. We studied different levels of heterogeneity: 2-level, 3-level and multi-level in terms of the node energy levels.

X. DISTRIBUTED EFFICIENT CLUSTERING APPROACH: DECA

DECA is an improved Distributed Efficient Clustering Approach [6]. The basic difference between the HEED and DECA is how the nodes take the decision and the score computation. The phases involved in DECA operation Start ClusteringReceive Clustering Message actual announcementFinalize Clustering.

XI. EFFICIENT ENERGY BASED MULTIPATH CLUSTER ROUTING PROTOCOL: EEMCRP

This minimizing the energy consumption it provides load balancing and increased throughput to the network. The factor which attain balance end to end delay, delivery rate and energy efficiency in WSNs.EMCRP achieves better data delivery ratio, improved network lifetime, less end to delay and energy consumption in terms of mobility, time, throughput, simulation time This provides the Energy Conservation without losing accuracy, Reliability based on Fault Tolerance and the Standard Quality of Service.This mainly evaluated the performance metrics of control overhead,throughput,packet delivery ratio. Past recent work shows more attention on this critical issue and, about designing different energy efficient routing protocols with different energy efficient mechanism.

XII. TREE BASED APPROACH

The anotherenergy efficient way of routing the data over the network is tree based approach. In this approach a hierarchical manner of aggregation points are formed which resembles the tree structure. The leaves are the source nodes and the root is the sink node. The data when travelling gets aggregated in the intermediate nodes itself. The most successful energy efficient routing protocol which follows the tree based approach was PEGASIS.

XIII. HYBRID ROUTING PROTOCOL: HRP

Hierarchical routing protocols are best known in regard to energy efficiency. By using a clustering technique hierarchical routing protocols greatly minimize energy consumed in collecting and disseminating data .In this protocol is used to decrease the probability of failure nodes and to prolong the time interval before the death of the first node (stability period) and increasing the lifetime in heterogeneous WSNs, which is crucial for many applications. HRP is a hierarchical protocol based on GPS (Global positioning system). Thehierarchical protocol is an extension of the Zone-based Hierarchical Link State.

Reactive routing is used in the destination zone. It reduced the network traffic and the consumption of bandwidth andalso increasedthe network performance.

Table 1.

S. no.	Protocols	Proved Result
1	MAC protocols	It focusesOnenergy consumption, latency, and reliability
2	ODSAA and ODAA protocol	Improve the data throughput
3	Data timed sending protocol(DTS)	Improve the load balancing of the network
4	Equalized Cluster Head Election Routing Protocol (ECHERP)	The performance evaluation of ECHERP is carried out through simulation tests, which evince the effectiveness of this protocol in terms of network energy efficiency
5	Position Responsive Routing Protocol" (PRRP).	CELRP initially shows good performance for average energy per packet compared to PRRP. However, after a short span of time PRRP shows a significant improvement in it with the increased number of data periods.
6	Clustering approaches	The purpose of balancing the load and prolonging the network lifetime.
7	Power-Efficient Gathering in Sensor Information Systems (PEGASIS)	This approach will distribute the energy load .It was represented by "limited battery capabilities
8	Directed Diffusion routing protocol	Provides Both security and energy efficiency together in wireless.The lifetime of the Network will be extended.
9	Low energy adaptive clustering hierarchy (LEACH)	This protocol allows us to space out the lifespan of the nodes, allowing it to do onlythe minimum work it needs to transmit data

XIV. CONCLUSION

So far various protocols have been discussed in this paper related how to improve the performance of battery for sensor nodes

Thedifferent types of protocol had provided improvement gains in Energy efficiency, Throughput, Delay, Bandwidth and Delivery Ratio. The feasibility of using the clustering technique and data aggregation needs to be tested in the same wireless sensor network.

REFERENCES

[1]. S. U. Hashmi, M. Rahman, H. T. Mouftah, and N. D. Georganas, "Reliability model for extending cluster lifetime using Backup Cluster Heads in cluster-based Wireless Sensor Networks," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2010 IEEE 6th International Conference on*, 2010, pp. 479-485.

[2]. Z. Miao, Y. Yuanyuan, and W. Cong, "Mobile Data Gathering with Load Balanced Clustering and Dual Data Uploading in Wireless Sensor Networks," *Mobile Computing, IEEE Transactions on*, vol. **14**, pp. 770-785, 2015.

[3]. J. Wang, W. Dong, Z. Cao, and Y. Liu, "On the delay performance in a large-scale wireless sensor network: Measurement, analysis, and implications," *IEEE/ACM Trans. Netw.*, vol. **23**, no. 1, pp. 186–197, Feb. 2015.

[4]. Rahul C. Shah and Jan M. Rabaey" Energy Aware Routing for Low Energy Ad Hoc Sensor Networks"F29601-99- 1- 0169 entitled, "Communication/Computation Pico nodes for Sensor Networks".

[5]. A. S. Wander, N. Gura, H. Eberle, V. Gupta, and S. C. Shantz, "Energy analysis of public-keycryptography for wireless sensor networks," in Third IEEE International Conference on Pervasive Computing and Communications (PERCOM'05). IEEE Computer Society Press, 2005, pp. 324-328.

[6]. Ankit Thakkar and Ketan Kotecha, "Cluster head election for energy and delay constraint application of wireless sensor network," *IEEE Sensors Journal*, vol. **14**, no. 8, August 2014.

[7]. Xi. Chen, Honglian Ma and Keqiu Li. A Geograhpy*based Heterogeneous Hierarchy routing Protocol Wireless Sensor Networks IEEE International Conference on High Performance computing and communications, 2008, pp. 767–774.

[8]. Lindsey, S.; Raghavendra, C. PEGASIS: Power-Efficient GAthering in Sensor Information Systems. In Proceedings of the IEEE Aerospace Conference, Los Angeles, MT, USA, 2002; pp. 1125–1130.

[9]. Miau Yu, Jason H.Li and renato Levy, "Mobility Resistant Clustering in Multi-Hop wireless Networks", *Journal of Networks*, Vol. **1**, No.1, May 2006.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 34-37(2018)

International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Survey on Image Forgery Detection Techniques

K. Ketzial Jebaseeli[1] and Dr. V.G. Rani[2]

[1]Research Scholar, Sri Ramakrishna College of Arts and Science for Women, Coimbatore (Tamilnadu), India

[2]Associate Professor, Sri Ramakrishna College of Arts and Science for Women, Coimbatore (Tamilnadu), India

ABSTRACT: Images are the means of communication. But in this day and age, an assortment of apparatus are obtainable that maneuvers the image. So the forged image has to be detected so that the image is not tainted and we can be sentient about the forged image and the real image. There are two types of image forgery detection copy move and image splicing, and diverse assails like smudges, blare, scaling, and many like this can take place. The outline of forgery detection techniques, the indispensable surge of how the forged image can be detected is presented in this paper. And to end with it is fulfilled with the proportional study with a quantity of parameters, merits and demerits.

Keywords: Forgery, copy-move, splicing.

I. INTRODUCTION

In our day to day digital world, digital images are the chief cause of information and they are the furthermost means of information pass on. As an indication for some events in the patio of edict images can be of use. The images were put out in TV news are time-honored as credentials for the straightforwardness of that news. Digital images are being used in numerous applications like military, therapeutic diagnosis, art piece, cinematography etc. Hence the digital illustration forensics emerges as hasty growing need of the public and so it is imperative to have an substantiate image. But in today's state of affairs the image forensics has befallen very unproblematic to stage-manage the digital images without leaving the discernible traces of exploitation with the assist of computers and accessibility of low-cost hardware and software tools[1]. So it becomes intricate for humans to sketch out these manipulations. As an upshot, the integrity and dependability of digital images are vanished. This amendment of images can be prepared for hiding some significant traces from an image, to revolutionize the details of an image so that erroneous information is put on the air. Consecutively to make out the veracity of the images we need to categorize any variation on the image

Fig. 1. Image forgery.

Fig. 2. Example of image forgery.

Passing on the example of image forgery, in the above figure 1 and 2, we can see that A is the real image in which they are disputing with the banners so that they can lend a hand to the black people to reside. Whereas in image B we can see that they are with the placards doing some encampment for not letting them to hang about and controlling colonization. We can evidently see the forged region stuck between the two images. The second image is the regional forgery that has been done using the hybrid wavelet transforms. According to some data, many journal-accepted documents contain figures with inapt and deceitful manipulations. Various methods encompassed to contradict tampering and forgery for ensuring image dependability. The foremost ambition of this paper is:

1. To momentarily commence what is image forgery.
2. To give outline of diverse techniques that are used to perceive forged spots.

The progression of creating counterfeit image has been immensely uncomplicated with the preamble of new-fangled and prevailing computer graphics editing software which are freely accessible as Photoshop, GIMP, and Corel Paint Shop. In the present day, this authoritative image processing software consent to people to amend photos and images opportunely and unperceivably. Now-a- days it creates a gigantic defy to

endorse the images. Sometimes it is complex to make out the edited constituency from the unique image. The detection of a forged image is obsessed by the necessitate of legitimacy and to maintain reliability of the image [2].

There are two approaches of forgery detection: 1. Active Approach 2. Passive approach. Figure 3, given below represents the flowchart of Active and Passive approach

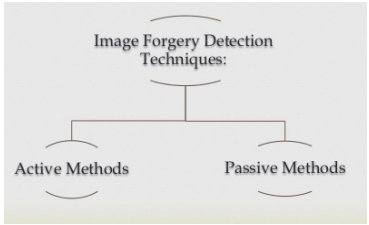


Fig. 3. Flowchart.

The active approach consists of two divisions watermarking and steganography. These two divisions re put into practice at the time of image attainment. An exceptional hardware accomplishment like digital signature or coding the image into unusual form is needed to blotch the endorsement of the digital image. The watermarking technique is used to put out of sight a mark or a message in a depiction in order to defend its copyright at the time of image attainment and to ensure the legitimacy his message is hauled out from the image and demonstrated with the original watermarks. Hiding the imperative message so that it is not tainted by any third party is called steganography.

The passive approach does not entail any preceding information about the image and it is reliant on the traces left on the image by diverse handing out of steps all through image manipulation. With the facilitate of diverse image forgery detection practices the forged area, location and the amount of forgery can be perceived. It comprises of copy move forgery detection and image splicing and they also help to detect the operations that occur, like rotation, scaling, blurring etc.

II. TYPES OF IMAGE FORGERY

A. Copy-Move (Cloning)

Copy-move image forgery is transpired by copying and pasting the content within the same image. An image where an expanse of an image is copied and moved to an additional part in the same image is meant to be as tampered image. There are two sub types of copy-move, that are copy move and copy create [4].

Copy move image: It is occurred by means of copying one area within an image and pasting it onto another. In the given figure 4, B is the original image and A is the tampered image. The region that is marked rectangles are tampered images that are not in the original image(B) This type of image comes under copy-move image.

A. tampered image B. original image

Fig. 4. Copy move cloning.

Copy create image: Copy create image forgery is created by making use of one or more distinct and diverse images. An assortment of parts are copied and pasted from dissimilar images and a forged image is shaped [4]. In Figure 5, the three distinct images are formed and represented as a single forged image. In the first image, one person is imaged, in the second image one, and from the third image all the three combined to form a solitary image that is forged. In simple words, the three images are combined and fused together.

Fig. 5. Copy Create Image [4].

B. Image Splicing

Image splicing is a process in which it crops and pastes regions from the same or images that are poles apart. Digital photomontage uses image splicing so that two images can be stuck together using tools like Photoshop [3]. Image splicing indicates the copying of one image and pasting it into another image. Passing on the figure 6, the left side comprises of two different images which are spliced to form a single merged image given towards the right side.

Fig. 6. Image Splicing.

III. TYPES OF ATTACKS

There are assorted forms of attacks that occur during forgery that is blurring, scaling, noise addition, cropping, compression, rotation, resizing, retouching, down sampling, in painting and so forth. Below are some of the attacks with example.

A. Blurring

Blurring happens when one image is copied or pasted into another image or due to various reasons, or at times it may happen despite the fact that on taking pictures the camera may not be tranquil or sometimes may be of moving objects be it a tree or a bird, blurring happens. This type of attack comes under blurring [5].

Fig. 7. Blurring.

The above given figure 7, is the distinguished image of a blurred image and a real image.

B. *Cropping*

It confiscates the outer part of an image to perk up the subject matter, framing or amend the aspect ratio. Depending on the use, this may be carried out on a physical photograph, artwork or film footage, or achieved digitally by means of image editing software. The practice is well-known to the film, broadcasting, photographic, graphic design as well as printing industries. The figure 8 given below is the distinct image of an animal that is cropped on the left and original image on the right.

Fig. 8. Cropping.

C. Retouching

Retouching can be carried out with paint, ink, piecing photos or negatives together in the darkroom, scratching, double exposure etc. Shifting the image with an object to boost up its quality is called as image retouching. Retouching is all about making infinitesimal changes and making the image good-looking. Image retouching can be done on product image, model image, creases etc. In the figure 9 given below is an example of an original image and a retouched image.

Fig. 9. Retouching.

D. Scaling

The process of resizing a digital image is referred as image scaling. It is a non-trivial process that entails a trade-off flanked by sharpness, smoothness and competence. When scaling a vector graphic image, the graphic archaics that formulate the image can be scaled using geometric transformations, with no trouncing of image excellence. When scaling a raster graphics image, a new image with a superior or inferior number of pixels must be spawned. In the case of decreasing the pixel number (scaling down) this typically results in a visible quality failure.

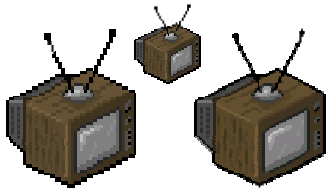


Fig.10. Scaled image.

From the viewpoint of digital signal processing, the scaling of raster graphics is a two-dimensional paradigm of sample-rate alteration, the adaptation of a signal that is standing apart from a sampling rate to an added signal [6]. The figure 10 given below indicates the scaling of an image into different sizes.

E. Inpainting

To assess the stoutness of the proposed technique against diverse attacks, we designed an experiment involving an in painting attack. Image in painting is an aggressively growing field of research for the reason that it can efficiently patch up the damaged or impassive regions in a visually conceivable way. All of the sample images were obtained from public online databases. In this try out, two widespread evaluation norms (CDR/FDR) were not pertinent because the copied regions of inpainted images are unidentified [7][8]. The figure 11 represented below is the inpainted image that is shaded in green.

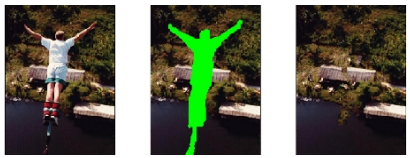


Fig. 11. Inpainted image.

IV. CONCLUSION

In this paper, we have discussed about image forgery detection, different categories of image forgery that come about and the attacks that occur due to forgery This paper is also comprised of Copy-move method and it is a widespread method for image forgery. It works devoid of any digital watermarks or signature information. Copy-move forgery, while remaining robust against actions aimed at concealing forgery, including slight rotation, slight scaling, JPEG compression, blurring, and brightness adjustment. This study, consequently, makes the most valued contribution to the field of multimedia forensics. Nevertheless, image amendment can be obscured by methods of greater erudition, such as great rotation, scaling, noise addition, inpainting or a permutation thereof. This makes the detection of copy-move forgery far more exigent. We are currently mounting on the methods to prevail over these limitations.

REFERENCES

[1]. P. Chi-Man, X. Yuan, and X. Bi, "Image Forgery Detection Using Adaptive Over-Segmentation and Feature Points Matching," *IEEE Transactions on Information Forensics and Security*, (2015).

[2]. H. Chen-Ming, J. Lee, and W. Chen, "An Efficient Detection Algorithm for Copy-Move Forgery," *Information Security (AsiaJCIS), 10th Asia Joint Conference on. IEEE,* (2015).

[3]. B. Khosro, and A. Kot, "Image splicing localization based on blur type inconsistency," Circuits and Systems (ISCAS), *IEEE International Symposium on. IEEE*, (2015).

[4]. A. Anoop, "Image forgery and its detection:A survey," *IEEE Sponsored 2J/d International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS), (2015).*

[5]. A. Edoardo, A. Bruno, and G. Mazzola, "Copy-move forgery detection by matching triangles of keypoints," *IEEE Transactions on Information Forensics and Security*, (2015).

[6]. M. Ansari, S. Ghrera and V. Tyagi, "Pixel-Based Image Forgery Detection: A Review," *IETE journal of education*, Vol. **55**, no 1, jan-jun, (2014).

[7]. A. Kaur, and R. Sharma, "Copy-move forgery detection using DCT and SIFT," *International Journal of Computer Applications*, **70**, no. 7 :30-34 (2013).

[8]. H. Jie, H. Zhang, Q. Gao, and H. Huang, "An improved lexicographical sort algorithm of copy-move forgery detection," *In Networking and Distributed Computing.*

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 38-43(2018)

International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Study and Review on Data Mining Based Email Fraud Detection Techniques

N. Geetha[1] and P. R. Kalaiyarasi[2]

[1]Assistant Professor in Computer Applications, Vellalar College for Women, Erode (Tamilnadu), India.
[2]Assistant Professor in Computer Applications, Vellalar College for Women, Erode, (Tamilnadu), India.

ABSTRACT: E-mail is one of the most widely used ways of written communication over the internet, and its traffic has increased exponentially with the advent of World Wide Web. The increase in email traffic comes also with an increase in the use of emails for illegitimate purpose. In this paper, we present a brief survey of the major research efforts on email mining. Phishing, Spamming, email bombing, threatening, cyber bullying, racial vilification, terrorist activities, child pornography and sexual harassment are common examples of e-mail abuses. So, there is a need for e-mail mining. Various methods and approaches were used by the scientists for classification of email messages in above categories. The focus of this review study is to summarize all existing email fraud detection techniques. In this paper we are presenting various techniques and approaches used by researchers for email fraud mining and subsequent classification.

Keywords: Email, Spam email detection, Suspicious email detection, Phishing email detection, Email authorship identification

I. INTRODUCTION

A network is consisting of actors and their social ties to each other. The social process is an important, hard to study, aspect of any software engineering effort. Nearly all communication is done via internet. Records of both communication and development activity are freely available. Email is one of the most popular forms of communication nowadays, mainly due to its efficiency, low cost, and compatibility of diversified types of information. In order to facilitate better usage of emails and explore business potentials in emailing, various data mining techniques have been applied on email data.

Email is one of the crucial aspects of web data communication. The increasing use of email has led to a lucrative business opportunity called spamming. A spam is an unwanted data that a web user receives in the form of email or messages. This spamming is actually done by sending unsolicited bulk messages to indiscriminate set of recipients for advertising purpose. These spams messages not only increases the network communication and memory space but can also be used for some attack. This attack can be used to destroy user's information or reveal his identity or data. In this paper we discuss some approaches for spam detection [10].

Email has been an efficient and popular communication mechanism as the number of Internet users increase. Therefore, email management is an important and growing problem for individuals and organizations because it is prone to misuse. The main objective here is to identify e-mails that contain suspicious contents indication future criminal activities. For example, if domain specific suspicious keywords (kill, attack, bomb, etc.) are found in an email message, it is classified as suspicious where as if non-suspicious indicators are present in an email; it is further classified as non-suspicious or may-be-suspicious.

Phishing emails are specialized class of illegitimate emails, which are intended to obtain useful information from the receiver of email. Phishing problem is believed to be a security and privacy concern. Phishing problem is considered to be the hard problem, due to the fact that an attacker can easily make the replicated website which may resemble to the legal bank of a user. Phishing emails are the emails which are planned to acquire crucial information from the receiver. The crucial information includes username, password, credit card details, bank account information, etc. These emails resemble to the emails from trustworthy websites. The emails contain such a text that the receivers immediately turn to respond the email by clicking on the links provided in the email or send the crucial information in reply [9].

Authorship Identification techniques are used to identify the most appropriate author from group of potential suspects of online messages and find evidences to support the conclusion. Cybercriminals make misuse of online communication for sending blackmail or a spam email and then attempt to hide their true identities to avoid detection. Authorship Identification of online messages is the contemporary research issue for identity tracing in cyber forensics.

This is highly interdisciplinary area as it takes advantage of machine learning, information retrieval, and natural language processing. In this paper, a study of recent techniques and automated approaches to attributing authorship of online messages is presented.

E-Mail Mining Fraud Detection Categories

- Spam email detection
- Suspicious email detection
- Phishing email detection
- Email authorship identification

II. SPAM EMAIL DETECTION

There are several approaches to identify incoming messages as spams are, Whitelist/Blacklist, Bayesian analysis, Mail header analysis, Keyword checking etc. some of them are defined below:

Whitelist/Blacklist: These approaches simply create a list. A whitelist is a list which includes the email addresses or entire domains which the user knows. An automatic white list management tool is also used by user that helps in automatically adding known addresses to the whitelist. A blacklist is the opposite of whitelist. In this list we add addresses that are harmful for users.

Mail Header Checking: This approach is very known approach. In this we simply consist of set of rules that we match with mail headers. If a mail header matches, then it triggers the server and return mails that have empty "From" field, that have too many digits in address that have different addresses in "To" field from same source etc.

Signatures: This approach is based on generating a signature having unique hash value for each spam message. The filters compare the value of previous stored values with incoming emails values. It is probably impossible for legitimate message having same value with spam message value stored earlier.

Bayesian Classifier: There are particular words used in spam emails and non spam emails.

Table 1: Advantages and Disadvantages of Spam Email Detection Techniques.

Approach	Advantage	Disadvantage
Whitelist/Blacklist	-Simplistic in nature	-Easily penetrated by spammer.
Signatures	-Low level of false positives	-Unable to identify spam until email reported as spam & its hash distributed.
Mail Header Checking	-Easily implemented	-High false positive rate and rejecting connections require additionalinformation/policies.
Bayesian Analysis	-State-of-the-art approach (wide – spread implementation).	-Rely on 'naive' Bayesian filtering (which assumes events occurred independent each other).

These words have particular probability of occurring in both emails. The filters that we used don't know these probabilities in advance; we must train them first so it can build them up. After training the word probabilities are used to compute the probability that an email having particular set of words in it belong to either spam or legitimate emails. Each particular word or only the most interesting words contribute to email's spam probability. This contribution is known as the posterior probability and is computed using Bayes' theorem. Then, the emails spam probability is computed all over the word in the emails. If this total value exceed over certain threshold then the filters will mark emails as spam.

III. SUSPICIOUS EMAIL DETECTION

The detection of criminal activities on E-mail needs text mining, machine learning and NLP techniques and methodologies to form and select pattern and knowledge from the E-mails. The main aim of this survey is to explore the existing literature, E-mail representation and classification techniques [12]. Classification of E-mails is a crucial issue. The execution performance of a classification algorithm in data mining is dependent upon quality of the data source. Insignificant and redundant features of data not only increase the cost of mining process, but also affect the quality of the result in some cases. Each algorithm has its own benefits and drawbacks as described in this section.

There are various classification techniques for e-mails.

A. Decision Tree

The decision tree restructures the manual categorization of training documents by discovering well-definednodes in a tree structure. In a decision tree structure it consists of two types of nodes one is internal node another one is external node. Internal nodes correspond to attributes selected by decision tree algorithm for making decision at specific level of hierarchy. The branches coming out from these internal nodes are the values of that attribute. The attribute at the top level of hierarchy in the tree has more power of classifying the instances of different classes[1].

B. Decision Rules Classification

A decision rules classification method uses the rule based assumption to classify E-mail to their annotated categories. The algorithms composed a set of rules that describes the profile for each category. During the classification process, it is not necessary that each rule existing in the rule set needs to be satisfied. In the case of handling a dataset for each class with a large number of features, heuristics implementation is prescribed to reduce the size of rules set without influencing the performance of the classification[2].

C. Naïve Bayes Algorithm

A simple probabilistic classifier named as Naïve Bayes classifier is based on applying Bayes' Theorem with strong independence assumptions.Naïve Bayes is a statistical analysis algorithm which works on numeric data.

D. Support Vector Machine (SVM)

Support vector machines (SVMs) are selective classification methods which can be recognized as more accurate. The SVM classification method is based on the Structural Risk Minimization from computational learning theory. The principle is to find a hypothesis to assurance the lowest true error. SVM has some impressive features for it has been considered as the state-of-art in the text classification tasks. SVM has been used as different classification tasks such as text classification, hand written digit detection etc.

E. Artificial Neural Network (ANN)

A neural network machine learning algorithm is an attempt to mimic the way real neural networks in the human brain work. Each neural network has an input layer and an output layer, with one or more hidden layers between them. Depending on the input to each neuron, they will pass their output on to the next layer of neurons. Propagating values forwards thorough the network like this makes values eventually reach the output layer, where predictions can be made. For document classification tasks, different kinds of ANN approaches have been introduced. Due to simplicity in implementation some researchers used Single layer perceptron, which contains only an input layer and an output layer. Inputs are fed directly to the outputs via weights in series. This approach is treated as the simplest kind of feed-forward network. The multi-layer perceptron which is more sophisticated consists of an input layer, one or more hidden layers, and an output layer in its structure, also widely implemented for classification tasks [6].

F. Genetic Algorithm based classification

Genetic Algorithms can identify and exploit regularities in the environment, and converges on solutions that were globally optimal. This method is very effective and widely used to find-out optimal or near optimal solutions to a wide variety of problems. Genetic algorithm does not impose any limitations required by traditional methods such as gradient descent search, random search etc. The Genetic Algorithm technique has many advantages over traditional non- linear solution techniques. However, both of these techniques do not always achieve an optimal solution. However, GA provides near optimal solution easily in comparison to other methods.

Table 2: Advantages and Disadvantages of Suspicious Email Detection Techniques

Approach	Advantage	Disadvantage
Decision Tree	- It is simple to build and it is an inductivealgorithm in which generated rules areeasily understandable by humans, and provide aconsolidatedperspective of the classification logic[3].	-The occurrence of an alternative tree, it over-fitsthe training data that further categorizes the trainingdata poorly but would classify the documents to beclassified superior[3].
Decision Rules Classification	-In classification task to construct local dictionary for each individual category in theimplementation of decision rules method.	-Decision rule process is the inconveniences to assign a document to exclusively. -The learning and updating of decision rule methods need extensive involvement of human experts to discover or update the rule sets. -When the number ofdistinguishing features are large, the decision rules method does not work properly.
Naïve Bayes Algorithm	-It Requires a small amount of training data to estimate the Parameters necessary for classification. -Naïve Bayes is a simple and fast classifier. -It works well with statistical representations such as bag-of-words.	-The Naive Bayes classifier requires a very large number of records to obtain good results. -Less accurate as compared to other classifiers on some datasets. -Optimal Bayes classifier computationally intractable.
Support Vector Machine	-It can work well in a very high dimensional feature space, it uses only a subset of original training to make decision boundary called support vectors and it is also suitable for non-linear separable data.	-It takes long training time. -It is difficult to understand the learned function. -It is not easy to incorporate domain knowledge.
Artificial	-It able to handle the	-All inputs have to

Neural Network	documents with high dimensional features and also handle the documents with noisy and contradictory data. -Neural network has been applied in document classification systems to improve efficiency. -Neural Network for document classification produces good results in complex domains and is suitable for both discrete and continuous data (especially better for the continuous domain) [7].	be translated into numeric inputs. -Slow training. -Learning might result in a local optimum.
Genetic Algorithm based Classification	-It does the encoding of the parameters, not the parameters itself; the search is more elaborative in a given amount of time. -GA is probabilistic in nature, it may yields different solutions on different set of simulations	-InGenetic Algorithm, the Fitness function must be chosen very carefully. If the fitness function is chosen poorly, then Genetic Algorithm may not be able to find an optimal solution to the problem, or may end up solving the wrong results. -Genetic Algorithms usesrandom parameterselection, hence it will not work well when thepopulation size is small and the rate of change is toohigh.

IV. PHISHING EMAIL DETECTION

Phishing attack is a kind of network attack which theft the identity of user's online and steals some useful information such as password or ATM and financial information. The phishing is classified into two categories deceptive phishing and malware-based phishing. In this, literature study of some of the previously work done to prevent network from phishing attack is described with their merits and demerits. Various anti-phishing techniques have been evolved to protect our website/ link and personal information against phishing attacks[11].

A. List Based Approach

This is possibly the most straightforward solution for anti-phishing. A white list contains URL's of known legitimate sites. Many current anti-phishing techniques rely on the combination of white list and blacklist. The representative blacklist/white list based systems include Phish Tank Site Checker, Google Safe Browsing, Fire Phish and Calling ID Link Advisor. This anti-phishing result would generally deploy similarly as toolbars or extension of web browsers should remind those clients if they would scan a sheltered websites. Blacklist undergo from a window of vulnerability between the time a phishing site is launched and the site's addition to the blacklist as it requires frequent updating which is the case for white list also [13].

B. Ant Colony Optimization

The Ant Colony System is based on the basic idea of an ant food searching system. This novel heuristic known as Ant Colony Optimization (ACO) has been found to be mutually vigorous and multipurpose in handling an extensive range of combinatorial optimization problems. The major suggestion of ACO is to model a predicament as the search for a least cost path in a graph. Artificial ants as if walk on this graph, gazing for cheaper paths. Each ant has a somewhat uncomplicated behavior accomplished of finding comparatively costlier paths. Cheaper pathways are found as the growing consequence of the universal cooperation among ants in the colony.

C. PhishZoo

It can detect current phishing sites if they look like legitimate sites by matching their content against a saved profile. In order to avoid detection, a phishing site must gaze fundamentally unique in relation to a genuine website. Our working assumption is that such different-looking sites have a better chance of catching users' attention about their phishiness. PhishZoo can be used to improve security by attackers to misuse client trust [14].

D. K-NearestNeighbor (k-NN)

This Classifier proposed for phishing email filtering. Using this classifier, the decision is made as follows: based on k-nearest training input, samples are chosen using a pre-defined similarity function; after that, the email x is labeled as belonging to the same class as the bulk among this set of k.

E. Fuzzy Logic

The fuzzy logic techniques presents more information to help risk managers successfully manage assessing and ranking website phishing risks than the existing qualitative approaches as the risks are quantified based on a amalgamation of historical data and practiced input [8].

F. Genetic Algorithm

Genetic algorithms can be used to develop simple rules for preventing phishing attacks. These rules are used to differentiate normal website from anomalous website. These anomalous websites refer to events with probability of phishing attacks. The objective of applying GA is to generate rules that match only the anomalous URLs of websites. These rules are tested on historical URLs and are used to filter new URLs to find suspicious phishing attacks.

Table 3: Advantages and Disadvantages of Phishing Email Detection Techniques.

Techniques	Advantages	Disadvantages
List Based Approach	- This approach is 100 % accurate on decision for blacklisting of website - This approach also produce less false positive rate - It also requires less computational cost and easy to use.	-It produce much memory overhead - If the websites are not in the list of blacklist then the accuracy is nil
Ant Colony Optimization	-This approach is accurate by determining the best rules or features -can be used in dynamic environment -retain memory of entire colony	- It enhance false negative rate as compare to other ones.
PhishZoo	-It can classify zero-day phishing and targeted attacks - This approach also able to detect new attack -Reduces false positive rate	-It less robust for detection of phishing - It requires matching image site
K-nearest neighbor	-It is much capable to achieve a true positive rate - Capable to achieve high accuracy	-huge number of feature - Higher cost - High memory requirement
Fuzzy Logic	-It requires less memory -its inference speed is also very high	- It is not 100% effective - It is complex to design
Genetic Algorithm	-It is better in classifying the email message as phishing mails -It produce less	-It requires more domain specific knowledge - They are not easy to handle it.

V. EMAIL AUTHORSHIP IDENTIFICATION

Email authorship identification is considered to be the task of identifying the most probable author of an email by analyzing the past emails of the suspected authors. The importance of writeprints is in order to prevent cybercrimes. Authors argued that writeprints are as important as fingerprints are for identifying the criminals in real life. The authors presented a write-print based model for mining frequent patterns in the emails in order uniquely identify the authors of emails. The Authorship Attribution Techniques is classified in two categories are statistical Analysis [5] and Machine learning techniques.

(I) Statistical Univariate Methods

A. Naive Bayes classifier

In this Classifier Learning and classification methods based on probability theory. Bayes theorem plays a critical role in probabilistic learning and classification. It uses prior probability of each category given no information about an item.

B. CUSUM statistics procedure

In statistical analysis the cusum called cumulative sum control chart, the CUSUM is a sequential Analysis technique used for monitoring change detection. As its name implies, CUSUM involves the calculation of a cumulative sum.

C. Cluster Analysis

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its purpose is to sort cases (people, things, events, etc) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters.

(II) Machine Learning Techniques

A. Feed-forward neural network

A feed forward neural network is an artificial neural network where connections between the units do *not* form a directed cycle. This is different from networks. The feed forward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

B. Radial basis function network

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks are used for function approximation, time series prediction, and system control.

C. Support Vector Machines

In machine learning, support vector machines (SVMs), also support vector networks are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier [4].

VI. CONCLUSION

This paper provides a review of machine learning approaches and classification. An analysis of feature selection methods and classification algorithms were presented. E-Mail classification required more works and efforts are required to improve the performance and accuracy of the process. New methods and solutions are required for useful knowledge from the increasing volume of electronics documents.

Spam emails are the biggest problem for the web data. This paper explored different approaches to deal with this problem. From all of these approaches no one can provide 100% result. Some of the approaches provide high false positive rates and false negative rates. There is very much scope for identifying mail as spam emails or legitimate mails for text as well as multimedia messages.

To detect suspicious criminal activities is a new active area of text mining. Automatic classification and analysis techniques are needed for detection of suspicious criminal activities on E-mails.

Phishing attack is one of the serious threats of network which stole the user's secret or confidential information. In this paper, we study different types of anti-phishing techniques. In future work, develop such technique which can detect such serious threat accurately and reduces the memory overhead together with decrease the false positive rate.

The problem of authorship attribution is explored well in the area of literature, newspapers etc but limited work has been done for the authorship identification of online messages like blogs, emails and chat.

REFERENCES

[1]. J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. **1**, no. 1, 1986, pp. 81-106, Kluwer Academic Publishers, Boston.

[2]. Russell Greiner, Jonathan Schaffer; AIxploratorium – Decision Trees, Department of Computing Science,University of Alberta,Edmonton,ABT6G2H1,Canada.2001.

[3]. URL:http://www.cs.ualberta.ca/~aixplore/ learning/ Decision Trees.

[4]. Chidanand Apte, Fred Damerau, Sholom M. Weiss.; "Towards Language Independent Automated Learning of Text Categorization Models", In Proceedings of the 17thAnnual International ACM-SIGIR Conference on Researchand Development in Information Retrieval, pp. 23-30.1994.

[5]. Vladimir N. Vapnik, "The Nature of StatisticalLearning Theory", Springer, NewYork. 1995.

[6]. Miguel E. Ruiz, PadminiSrinivasan; "Automatic Text Categorization Using Neural Network", In Proceedings of the 8th ASIS SIG/CR Workshop on Classification Research, pp. 59-72. 1998.

[7]. Bo Yu, Zong-ben Xu, Cheng-huaLi, "Latent semantic analysis for text categorization using neural network",

[8]. Knowledge-Based Systems 21- pp. 900–904, 2008.

[9]. Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them".

[10]. Send mail Home Page, http://www.sendmail.org, Accessed 01, July 2004.

[11]. Jagruti Patel, Sheetal Mehta, (2015). "A Literature Review On Phishing Email Detection Using Data Mining", *International Journal Of Engineering Sciences & Research Technology*, **4**(3): March, 2015 ISSN: 2277-9655.

[12]. B. Ross, C. Jackson, and N. Miyake, "Stronger Password Authentication Using Browser Extensions", In: Proc. of the 14th Usenix Security Symposium, 2005, pp.2-16.

[13]. E. Kirda, and C. Kruegel, "Protecting Users against Phishing Attacks with AntiPhish", *In: Proc. of the 29th Annual International Computer Software and Applications Conference*, 2005, pp.521-534.

[14]. A. P. E. Rosiello, and E. Kirda, et al, "A Layout-Similarity-Based Approach for Detecting Phishing Pages", *In: Proc. of the third International Conference on Security and Privacy in Communications Networks*, 2007, pp.454-463.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 44-48(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Study of Attacks, Security Mechanisms in Wireless Sensor Networks

K. Sutha[1] and S. Srividhya[2]
[1]Research Scholar, Department of Computer Science,
Sri Ramakrishna College of Arts & Science for Women, Coimbatore (Tamilnadu), India
[2]Assistant Professor, Department of Computer Science,
Sri Ramakrishna College of Arts & Science for Women, Coimbatore (Tamilnadu), India

ABSTRACT: This paper presents the review of the existing privacy techniques in Wireless Sensor Network. (WSN) can be used to monitor either physical or environmental conditions like temperature, sound pressure etc. Privacy of individuals is of very high demand as the popularity of wireless sensor network increase. This paper presents two main parts of WSN that are Types of Attacks and Security Mechanisms. This paper aims to study of security goals, attacks, security mechanisms to handle privacy issues.

I. INTRODUCTION

Most sensor networks actively monitor their surroundings, and its often easy to deduce information other than the data monitored. Security is critical for such networks deployed in hostile environments. Wireless sensor networks have many applications in military, homeland security and other areas. These factors demands Security mechanism for sensor networks at design time to make sure operation safety, secrecy of sensitive data, and privacy for People those who are using sensor networks.One of the major challenges in wireless sensor networks face today is security. While the deployment of sensor nodes in an unattended environment makes the networks vulnerable to a variety of potential attacks, the inherent power and memory limitations of sensor nodes makes conventional security solutions unfeasible. Hence security is a crucial issue. Many techniques have been implemented for the privacy preservation in WSN such as; cryptographic security method, k-anonymity . But these techniques were used to protect the information when it flows from one node to another node. A large number of attacks have been possible in WSN. This includes Sybil attacks, Traffic analysis attacks, physical attacks, Dos attacks. This paper includes a survey of attacks and security mechanism to handle attacks and challenges in Wireless Sensor Networks.

Some important security dependent applications include:

- **Disasters**: In many disaster scenarios, especially those encouraged by terrorist activities, it may be necessary to protect the location of casualties from unauthorized disclosure.
- **Public Safety**: In applications where chemical, biological or other environmental threats are monitored, it is vital that the availability of the network is never threatened. Attacks causing false alarms may lead to panic responses or even worse total disregard for the signals.
- **Home Healthcare**: In such applications, privacy protection is essential. Only authorized users should be able to query and monitor the network. The major contribution of this paper includes classification of security attacks, security mechanisms and challenges in Wireless Sensor Networks [2].

II. SECURITY GOALS FOR WIRELESS SENSOR NETWORKS

A sensor network is a special type of Ad-hoc network. So it shares some common information or property as computer networks. There are some security mechanisms and requirements to implemented in WSN. Those requirements are considered as security goals. Security goals are splitted in to two types. They are Primary and Secondary goals. The primary goals are known as standard security goals such as Confidentiality, Integrity, Authentication and Availability (CIAA). The secondary goals are Data Freshness, Self Organization, Time Synchronization and Secure Localization [4].

PRIMARY GOALS

A. Data Confidenteiality

Data confidentiality is the most important issue in network security. In sensor networks, the confidentiality relates to the following: A sensor network should not leak sensor readings to its neighbors. Confidentiality is the ability to conceal messages from a passive attacker so that any message communicated via the sensor network remains confidential.

B. Data Integrity

Data integrity in sensor networks is needed to ensure the reliability of the data and refers to the ability to confirm that a message has not been tampered, altered or damaged. The integrity of the network will be in trouble when:

- A malicious node present in the network injects false data.
- Unstable conditions due to wireless channel cause damage or loss of data.

C. Data Authentication

Authenticating the source of message; -Entity Authentication: authenticating the user / node / base - station is indeed the entity whom it claims to be. Data authentication verifies the identity of the senders and receivers. Data authentication is achieved through symmetric or asymmetric mechanisms where sending and receiving nodes share secret keys. it is extremely challenging to ensure authentication.

D. Data Availability

Availability determines whether a node has the ability to use the resources and whether the network is available for the messages to communicate. ensuring desired service may be available whenever required. However, failure of the base station or cluster leader's availability will eventually threaten the entire sensor network. Thus availability is of primary importance for maintaining an operational network.

SECONDARY GOALS

A. Data Freshness

Freshness, which implies that the data is recent and ensures that no adversary can replay old messages Moreover, as new sensors are deployed and old sensors fail, we suggest that forward and backward secrecy should also be considered[8].

B. Self Organization

A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self-organizing and self-healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network.

C. Time Synchronization

Most sensor network applications rely on some form of time synchronization. Furthermore, sensors may wish to compute the end-to-end delay of a packet as it travels between two pairwise sensors. A more collaborative sensor network may require group synchronization for tracking applications.

D. Secure Localization

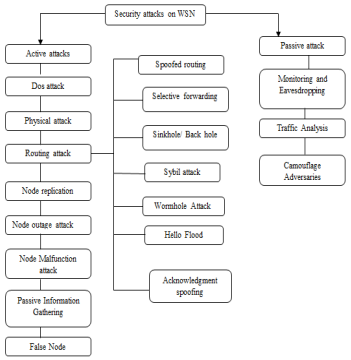
Often, the utility of a sensor network will rely on its ability to accurately and automatically locate each sensor in the network. A sensor network designed to locate faults will need accurate location information in order to pinpoint the location of a fault. Unfortunately, an attacker can easily manipulate non secured location information by reporting false signal strengths, replaying signals [6].

III. ATTACKS ON SENSOR NETWORKS

Attacks are divided into two types one is Active attack and another one is Passive attack.

A. Active Attack

In active attacks an adversary monitor**,** listens and introduces malicoious node, steal or modify message content, or break security mechanism.



Some of active attacks on WSN are:

Denial of Service Attack(DOS Attack). Denial of Service (DoS) is produced by the unintentional failure of nodes or malicious action. A simplest DoS attack on wireless sensor network is simply to jam a node or set of nodes. Jamming, in this case, is simply the transmission of a radio signal that interferes with the radio frequencies being used by the sensor network.

Physical Attack. Sensor networks typically operate in outdoor environments. Due to its unattended and distributed nature it is highly susceptible to physical attacks. For instance, attackers can extract cryptographic secrets, tamper with the associated circuitry, modify programming in the sensors, or replace them with malicious sensors under the control of the attacker [9].

Routing Attack. The attacks which act on the network layer are called routing attacks.

a) Spoofed routing information. In sensor network, sensor nodes takes some values and send it to the sink or base station .While routing this information to the base station or sink an attacker may alter or spoof that routing information to disrupt traffic in network

b) Selective packet forwarding. In wireless sensor network it is assumed that all nodes will accurately forward the entire received packet but in this attack an attacker creates malicious node which do not forward the entire message they received.

c) Sinkhole/Black hole Attack. In this type of attack, an attacker makes a compromised node look more attractive to neighbours nodes in sensor network with high resources capability. The result is that all surrounding nodes route their data a through this compromised node and when data routes through this compromised node it is able to do anything with the packet.

d) Sybil Attack. In wireless sensor network every sensor node might need to work together to accomplish a task. But in this attack a malicious node will appears as a set of nodes using the identities of other legitimate

nodes and affect routing mechanism, distributed storage, data aggregation and send incorrect information to the network.

e) Wormhole. In this attack, attacker records the packet (or bits) at one location in the network and retransmits those packets to another location

f) Hello flood. This is simplest attack in wireless sensor network in which an attacker broadcast "HELLO" packet as a weapon with high transmission to convince sensor nodes which are dispersed in large area within the wireless sensor network

g) Acknowledgement spoofing . Routing algorithms used in sensor networks sometimes require acknowledgments to be used [6] [7]. In this attack, an attacker may overhear packet transmitted from its neighbour nodes and spoof the acknowledgement or send false information (like send information that a node is alive when in fact it is dead) about nodes to the sending node.

Node Replication Attack. In this attack an attacker add a sensor node to existing sensor network with copying the node ID of existing sensor node. This attack reduces the network's performances by using packet corruption or misroute the packet.

Node Outage Attack. Node outage is the situation that occurs when a node stops its function. In the case where a cluster leader stops functioning, the sensor network protocols should be robust enough to mitigate the effects of node outages by providing an alternate route.

Node Malfuntion Attack. A malfunctioning node will generate inaccurate data that could expose the integrity of sensor network especially if it is a data aggregating node such as a cluster leader.

Passive Information Gathering

An intruder with an appropriately powerful receiver and well-designed antenna can easily pick off the data stream. Interception of the messages containing the physical locations of sensor nodes allows an attacker to locate the nodes and destroy them. Besides the locations of sensor nodes, an adversary can observe the application specific content of messages including message IDs, timestamps and other fields.

False Node

A false node attack is an attack in which attacker add a malicious node in wireless sensor network that feeds false data or prevents the passage of accurate data.

B) Passive Attack. The monitoring and listening of the communication channel by unauthorized attackers are known as passive attack. The Attacks against privacy is passive in nature [10].

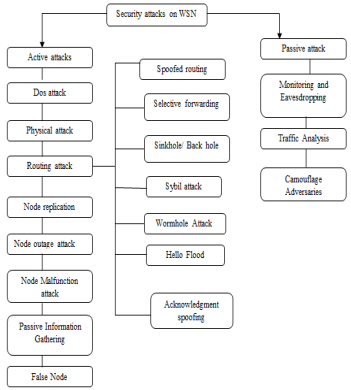
1) Monitor and Eavesdropping: This is the most common attack to privacy. By snooping to the data, the adversary could easily discover the communication contents. When the traffic conveys the control information about the sensor network configuration, which contains potentially more detailed information than accessible through the location server, the eavesdropping can act effectively against the privacy protection.

2) Traffic Analysis: Traffic analysis is a passive attack on wireless sensor network on privacy. In this attack an adversary analyze traffic that can identify activities in a wireless sensor network and also identify some sensor nodes that plays an important role in wireless sensor network. This attack may create Denial of Service (Dos) attack and also attack on those node which plays an important role.

3) Camouflage Adversaries. It is also a passive attack on wireless sensor network on privacy. In this attack an adversary either inserts a sensor node or compromises a sensor node in wireless sensor network. This camouflaged node attracts the packets from other nodes and may misroute those packet where privacy analysis perform.

IV. SECURITY MECHANISMS

The main motive behinds the security mechanisms is to detect, prevent, and recover from the security attacks to protect from different kind of security attacks. Based on this we can divides the security mechanisms into two parts.



A. Low Level Security Mechanism

1) Robustness to communication denial of service. An attacker attempts to disrupt the operations of the networks. Means, if a node have the huge power and broadcast a high energy signals. If the broadcasted node have highly transmission rate then the entire communication system channel could be jammed.

2) Secure routing. To Route the packet, is a crucial service for enabling the communication in sensor network. The Routing protocols we have till now, unfortunately suffer from many security vulnerabilities. An attacker can be launch the DoS (Denial of Services) attack on the routing protocol, preventing communication just by broadcasting a packet with high

transmission rate and could be jammed the entire communication network. An attacker may inject malicious routing information into the network, resulting in routing inconsistencies. We can protect the communication network from above attacks by using authentication.

3) Resilience to Node Capture. Resiliency against node capture is one of the most challenging issues in sensor networks. In most of the application, sensor nodes are to be placed on many locations that are easily accessible to the attackers. By this attacker can capture the sensor node and extract the cryptographic secrets, change their programs and may replace them with malicious nodes under the control of the attacker.

4) Key establishment and trust setup. The establishment of cryptographic keys is the primary requirement to setting up the sensor network. As we know, the sensors have the limited computational power and the public key cryptography primitives are too expensive. The way of communication of sensor network differ from traditional networks, sensor networks needs to establish a key and set up with their neighbors. The disadvantage of this mechanism is that the attacker may compromise and most of the node could be reconstruct the complete key pool and easily break the security scheme.

5) Secrecy and authentication. The sensor network application needs to protect against modification, eavesdropping and alteration of packet. To overcome with this, Cryptography is the standard defense approach. In the contest of cryptography, we can set up a centralized server which provides the unique id with key to each sensor node of the wireless network and whenever any sensor node wants to transmit the data over the network, first send to the server and then server will transmit the packet to the appropriate node. But problem of this approach is that if the server fails then the entire network leads to be failed. The earliest sensor networks are likely to use link layer cryptography, because this approach provides the greatest ease of deployment among currently available network cryptographic approaches [6].

6) Privacy. Like other networks, wireless sensor networks have also force privacy concerns. Initially the sensor networks are deployed for legitimate purpose might subsequently be used in unanticipated ways. Providing awareness of the presence of sensor node sand data acquisition is particularly important.

B. High Level Security Mechanisms

1) Secure Group management. Every node in wireless sensor network has the limited computational and communication capabilities. All the activities in a network like data aggregation and analysis of data can be performed by the group of nodes. In actual the nodes as comparing of the group may change continuously and quickly. The key services are also be performed by the group. That's why the secure protocol for group of nodes are to be needed, to securely admit new group members and provide the secure communication to newly admitted member within the group. The outcome of group key is transmitted to the base station. It is ensured that the output is authenticated and coming from authorized group. The group key must be managed by the group members and protects from outside member (nodes) or unauthorized nodes [7].

2) Secure Data Aggregation. The fine grain sensing is the advantage over the wireless sensor network which is provided by the dense sets of nodes. The data sensed by the different nodes in the network must be aggregated to avoid the huge amount of traffic back to the base station. Suppose a system may average the temperature of the geographic region, combine sensor values to compute the velocity and location of s moving object, or aggregate the data to avoid the false alarms in real-world event detection.

V. CONCLUSION

In this paper we surveyed on existing attacks on wireless sensor network. Also covered the security mechanism in WSN. In Security attack we described all the attack which is passive or active. We described some security goals to know how to handle and implement goals for security. And in security mechanism part we described how to detect prevent and recover from security attacks. This paper has proposed a set of guidelines to build privacy-related models in WSNs.

REFERENCES

[1]. Adrian Perrig, John Stankovic, David Wagner, "Security in Wireless Sensor Networks" Communications of the ACM, Page53-57, year 2004.

[2]. Al-Sakib Khan Pathan, Hyung-Woo Lee, ChoongSeon Hong, "Security in Wireless Sensor Networks: Issues and Challenges", Page1043-1045, year 2006.

[3]. Dr. G. Padmavathi, Prof and Head, Dept. of Computer Science, Avinashilingam University for Women, A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks year 2009.

[4]. Dr. Banta Singh Jangra, Vijeta Kumawat A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks year2015.

[5]. Abdul Wahid Pavan Kumar PG Student National Institute of Technology Patna National Institute of Technology Patna year 2012. "A Survey On Attacks, Challenges and Security Mechanisms In Wireless Sensor Network ".

[6]. Rajkumar Assistant Professor, Dept. of ISE, Sambhram Institute of Technology Bangalore Sunitha K R Lecturer, Dept. of ISE, Sambhram Institute of Technology Bangalore "A Survey on Security Attacks in Wireless Sensor Network " year 2012.

[7]. J. M. de Fuentes, L. González-Manzano, and O. Mirzaei Carlos III University of Madrid, Avenida de la Universidad 30, Legan´es, 28911Madrid, Spain Privacy Models in Wireless Sensor Networks: A Survey Volume 2016, Article ID 4082084.

[8]. Prakhar Gupta M.tech Scholar Department Of CSE, Maulana Azad National Institute of Technology Bhopal, Madhya Pradesh, India 'Privacy preservation for WSN: A Survey' Volume **48**– No.3, June 2012.

[9]. Pavitha N, Santosh N. ShelkeDepartment of Computer Engineering Sinhgad Academy of Engineering Pune, Maharashtra, India "Location Privacy in Wireless Sensor Networks" Volume **5**, Issue 5, MAY 2015 Pages 584-588.

[10]. S. SaravananResearch Scholar, Assistant Professor, Department of Computer Science, Government Arts College, Ariyalur, Tamilnadu, India "Location Privacy Based Security Enhancement In Wireless Sensor Network Using LFPM And PPM" Volume **11**, Number 6 (2016) pp 3936-3939.

International Journal of Theoretical & Applied Sciences, Special Issue 10(1a): 49-52(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

An Efficient TEEN Routing Protocol with DIJKSTRA Algorithm in Wireless sensor network

D. Barathi
Assistant Professor, Department of Computer Science,
Dr. R.V. Arts and Science College, Karamadai, Coimbatore (Tamilnadu), India

ABSTRACT: In dynamic wireless network connection based path failure problems and malicious packets declining are important factors for packet drops in multi-hop wireless ad-hoc network. The packet losses time cases are corresponding to the network channel allocation defect rate is based on observing the packet drop rate does not reached acceptable reliability. To improve the reliability, in this paper proposed a novel TEEN (Threshold Sensitive Energy Efficient Sensor Network Protocol) Routing Protocol with DIJKSTRA Algorithm method to develop the relationships between link failures. Meanwhile the proposed system to find the correlations, extend a DIJKSTRA algorithm based searching model framework design that allows to validates the reliability of the packet drops information by mobile nodes. Through the simulations, the system verifies and attains extensively enhanced learning accuracy across the existing techniques such as homomorphism linear authenticator based detection.

Keywords: TEEN, secure routing protocol, DIJKSTRA, Cluster.

I. INTRODUCTION

Mobile Ad Hoc Network (MANET) can be described as an autonomous collection of mobile nodes (users) that communicate over relatively low capacity wireless links, without a centralized infrastructure. In these networks, nodal mobility and the wireless communication links may lead to dynamically changing and highly unpredictable topologies. All network functions such as routing, multi-hop packet delivery and mobility management have to be performed by the member nodes themselves, either individually or collectively. So, network performance becomes highly dependent on collaboration of all member nodes. MANETs find applications in diverse fields ranging from low-power military wireless sensor networks to large-scale civilian applications, and emergency search/rescue operations.

There are good reasons why nodes in a mobile ad hoc network, that lacks the networking infrastructure which has been deployed through the investment of a telecommunications corporation, would prefer not to cooperate. When nodes do cooperate, they form the necessary ad hoc infrastructure that makes multi-hop communication achievable, allowing traffic from a node to reach destinations that would either require a significant amount of transmission energy using single hop communication, or simply not be possible without routing the traffic through other nodes. However, this means that nodes must be willing to forward traffic for other nodes, and in this way expend energy without receiving any direct gain from doing so. If a node only considers its own short-term utility, then it may not choose to participate within the network.

A malicious node that is part of the route can exploit its knowledge of the network protocol and the communication context to launch an insider attack—an attack that is intermittent, but can achieve the same performance degradation effect as a persistent attack at a much lower risk of being detected. Specifically, the malicious node may evaluate the importance of various packets, and then drop the small amounts that are deemed highly critical to the operation of the network. For example, in a frequency-hopping network, these could be the packets that convey frequency hopping sequences for network-wide frequency-hopping synchronization; in an ad hoc cognitive radio network, they could be the packets that carry the idle channel lists (i.e., white spaces) that are used to establish a network-wide control channel.

Current standards for wireless data communications such as IEEE802.11 and Bluetooth are easy targets of denial of service attacks. For example, the physical layers of IEEE802.11 and IEEE802.11b do not have any error-correction scheme. If an attacker sends a strong jamming signal of duration one bit/symbol it will make the CRC computation wrong. Therefore the whole packet will be lost. If we assume that this wireless link is used to transmit an IP data packet (usually 12000 bits long),the energy ratio between a jammer and user can be of the order of 1/10000 (which is equivalent to 40 dB gain for the jammer). Other wireless data standards that make use of error-correction codes can also be easily defeated. The reason is that current systems are designed to resist to non-malicious interference and noise. Even robust wireless links designed to resist jamming do not fully take into account the data aspect of the communication.

The rest of this paper is organized as follows. In Section 2 review the existing related work. The proposed models and descriptions are described in Section 3. Finally conclude the paper in Section 4.

II. RELATED WORK

In [1] authors discussed a security breach can, for example, be a network-level denial-of-service (DoS) attack, passive eavesdropping, or physical layer jamming to degrade communication channels. In a multi-hop network, an intruder node can degrade communication quality by simply dropping packets that are meant to be relayed (forwarded). The network could then misinterpret the cause of packet loss as congestion instead of malicious activity. The authors suggested that traffic transmission patterns be selected to facilitate verification by a receiver. Such traffic patterns are used in concert with suboptimal MAC that preserves the statistical regularity from hop to hop.

In [2] authors studied about the data delivery mechanisms that can with high probability circumvent black holes formed by these attacks. To argue that classic multipath routing approaches are vulnerable to such attacks, mainly due to their deterministic nature. So once the adversary acquires the routing algorithm, it can compute the same routes known to the source, hence, making all information sent over these routes vulnerable to its attacks. Authors developed mechanisms that generate randomized multipath routes. Under our designs, the routes taken by the "shares" of different packets change over time. So even if the routing algorithm becomes known to the adversary, the adversary still cannot pinpoint the routes traversed by each packet. Besides randomness, the generated routes are also highly dispersive and energy efficient, making them quite capable of circumventing black holes.

In [3] authors illustrated about the data integrity protection in Cloud Computing a very challenging and potentially formidable task, especially for users with constrained computing resources and capabilities. Thus, enabling public audit ability for cloud data storage security is of critical importance so that users can resort to an external audit party to check the integrity of outsourced data when needed.

To securely introduce an effective third party auditor (TPA), the following two fundamental requirements have to be met: 1) TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user; 2) The third party auditing process should bring in no new vulnerabilities towards user data privacy.

In [4] authors proposed four different jamming attack models that can be used by an adversary to disable the operation of a wireless network, and evaluate their effectiveness in terms of how each method affects the ability of a wireless node to send and receive packets. To discussed different measurements that serve as the basis for detecting a jamming attack, and explore scenarios where each measurement by itself is not enough to reliably classify the presence of a jamming attack. In particular, to observe that signal strength and carrier sensing time are unable to conclusively detect the presence of a jammer. Further, we observe that although by using packet delivery ratio we may differentiate between congested and jammed scenarios, we are nonetheless unable to conclude whether poor link utility is due to jamming or the mobility of nodes.

In [5] authors proposed a new routing service named best-effort fault-tolerant routing (BFTR). The design goal of BFTR is to provide packet routing service with high delivery ratio and low overhead in presence of misbehaving nodes. Instead of judging whether a path is good or bad, i.e., whether it contains any misbehaving node, BFTR evaluates the routing feasibility of a path by its end-to-end performance (e.g. packet delivery ratio and delay). By continuously observing the routing performance, BFTR dynamically routes packets via the most feasible path.

III. PROPOSED METHODOLOGY

The proposed method presents a precise algorithm for detecting selective packet drops made by insider attackers. The proposed methodology also provides a straightforward and visibly verifiable assessment statistics as a proof to support the detection decision. The methodology consists of four steps: (i) Network framework Configuration (ii) Graph Generation; (iii) Decision Tree formation; (iv) TEEN based Dijikstra algorithm. The proposed system architecture diagram as in Fig. 1.

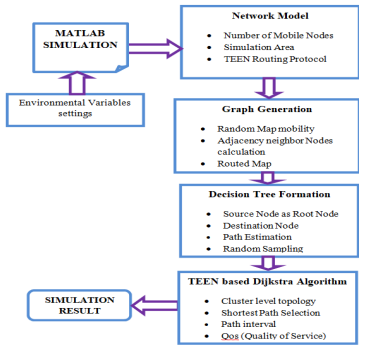


Fig. 1. Proposed System architecture.

A. Network Model

The mathematical model for TEEN based Dijkstra searching algorithm in wireless sensor network process. Let X be an number Nodes with area size = 1. where X_i = (i = 1,2, …, *n*).In the network model is created *N* number of nodes in the simulations area. As the amount of nodes in the wireless network is increased, the dimension of the simulation region is also increased so that a reliable node weights are maintained. Every node is moved according to the way of random mobility model.

The equation (1) is to set the nodes deployment in the simulation area is defined as follows:

$$X = area * rand(N, 2) \quad eqn.(1)$$

Where *X* is the Nodes, *N* is the Number of nodes has to deploy, area is the region size (i.e, Simulation area)

B. Graph Generation

The graph generation is the random map mobility routing process mechanism is initiated whenever a node needs to send a Control message to target node which is not in the transmit range consequently it must get a route to that corresponding node by initiation the Route detection process. This model usually the sender must first finds the routes in its cache if there is no pathway it precedes as follows: (a) It generates a route request packets having its locate the path and the address of the terminal node then it transmit the HELLO message packet to all its neighbors in broadcast manner. (b) Each neighbors receiving this packets request check with its cache is to find an final route to destination path to be returned turn around to the sender or else it re-transmit the same message information to all its connected neighbors after adding its location to the header of the request and learns from this message to be added to its routing cache. The mobile node has previously delighted this route request it disregards the new received request by verifying its sequence number since each route request is identified by a unique sequence number. (c) The similar procedure is executed by each nearby node until the route requests arrive to destination which adds its path at the end of the description and sends a route reply.

For Graph generation adjacency Distance Function is expressed as,

$$dist(i,j) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} X(i) - X(j) \quad eqn.(2)$$

Where, *X* is the Nodes, *i&j* is the nodes location.

C. Decision tree formation

The decision tree formation procedure is based on the shortest link path method is generally power saving optimized method. So individual metrics are considered and energy weight is allocated to the each path. The connecting the link end-to-end mobile nodes, there typically exists more than single way path. In the possible transmit node combinations, there will be comparatively energy best routes that achieve the lower cost based on the nodes' energy force and broadcast loss of the paths. A simple multi model wireless sensor network, with the transmit node set *T* between the source and terminal, and the instant neighbor set *T** for every node. Here exists an energy efficient route, for example, the route with dispatch nodes. The links with low broadcast power loss and mobile nodes with advanced remaining energy capacity are preferred. The problem is easy to minimize the power devoted during communication and exploit the battery energy of the nearby node to be used that is to minimize:

$$E(X(i)) = X(i).E - E_{Tx} + E_{mp}(4 * ch(dist)) \quad eqn.(3)$$

Where, *X* is the Nodes assignment, E_{Tx} is the Source node location; E_{mp} is the Energy Multipoint among the clusters; 4- represents 4- neighbor nodes; *ch(dist)* is the cluster head distance measures

D. TEEN Routing Protocol with DIJKSTRA Algorithm

TEEN is other hierarchical protocol for reactive networks that responds immediately to changes in the relevant parameters. In this protocol a clusters head (CH) sends a hard threshold value and a soft one. The nodes sense their environment continuously. The first time a parameter from the attribute set reaches its hard threshold value, the node switches on its transmitter and sends its data. The nodes then transmits data in the current cluster period if the following conditions are true: the current value of the sensed attribute is greater than the hard threshold, and the current value of the sensed attribute differs from sensed value by an amount equal to or greater than the soft threshold. Both strategy looks to reduce energy spend transmitting messages.

TEEN is appropriate for time basic applications and is additionally entirely effective as far as vitality utilization and reaction time. It additionally permits the client to control the vitality utilization and precision to suit the application.

Cluster Head Formation of Euclidean Distance formula as,

$$Ch(dist(X(i,j))) = \sqrt{X(i)_{xd} - X(i+1)_{xd}{}^2 + X(i)_{yd} - X(i+1)_{yd}{}^2} \; eqn.(4)$$

Where, *X* is the Nodes assignment, *xd*is the nodes x-coordinates, *yd* is the nodes y- coordinates.

Algorithm: TEEN Routing Protocol with DIJKSTRA Algorithm – Packet sending in Wireless Sensor Network

Step 1: Read the Number of nodes in the routing topology $X = \{x_1, x_2, \ldots, x_n\}$

Step 2: Analyzing the Neighbor nodes routing process and set the link to the nearest nodes.

Step 3: Initialize the Root node (i.e., Source Node) to generate the decision tree formation among the nodes.

Step 4: Set the Initial Energy value (0.1 joules) to all nodes and initialize Hard threshold and soft threshold value for learning the average energy statistics.

Step 5: Set the Cluster Head formation to group the nodes using Euclidean distance method.

Step 6: Set the Destination Node for packet sending process.

Step 7: To find the Shortest path among Source to Destination point Using TEEN based Dijkstra algorithm

a) Determine the Number of Nodes ids.
b) To calculate the number of nodes mapping using sparse representation on to Non-zero nodes active in the cluster head.
c) To find the shortest path nodes according to the nearest distance clusters to the destination node.
d) Calculate the Execution Time: Elapsed Time = (Finishing time – Initializing Time)
e) Calculate the Energy Loss as E_{Loss}= $E(X(i)) = X(i).E - E_{Tx} + E_{mp}(4 * c(dist))$

IV. CONCLUSION

In this paper needs to done by executing a wireless sensor network with correspondence convention utilizing TEEN (Threshold sensitive Energy Efficient sensor Network protocol) Routing Protocol with DIJKSTRA Algorithm. The Dijkstra searching algorithm manages the paths without packet dropping among the source to terminal nodes during the packet transmitting. The proposed technique using an extensive approach outperforms of dynamic TEEN routing algorithm optimization technique in terms of energy level and bandwidth discovery.

In future work, we intend to enhance the TEEN protocol to develop the experimental methods for non-linear optimization to control the growth of path of the result data.

REFERENCES

[1]. R. Rao and G. Kesidis, "Detecting malicious packet dropping using statistically regular traffic patterns in multihop wireless networks that are not bandwidth limited," in *Proc. IEEE GLOBECOM Conf.*, 2003, pp. 2957–2961.

[2]. T. Shu, M. Krunz, and S. Liu, "Secure data collection in wireless sensor networks using randomized dispersive routes," *IEEE Trans. Mobile Comput.*, vol. **9**, no. 7, pp. 941–954, Jul. 2010.

[3]. C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *Proc. IEEE INFOCOM Conf.*, Mar. 2010, pp. 1–9.

[4]. W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proc. ACM MobiHoc Conf.*, 2005, pp. 46–57.

[5]. Y. Xue and K. Nahrstedt, "Providing fault-tolerant ad-hoc routing service in adversarial environments," Wireless Pers. Commun., Special Issue Secur. *Next Generation Commun.*, vol. **29**, no. 3, pp. 367–388, 2004.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 53-57(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

An Overview of Risk Management Approach, Tools and Technology

K.K. Nivethithaa[1] and Dr. V. Krishnapriya[2]
[1]Research Scholar, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women, Coimbatore (Tamilnadu), India
[2]Head of the Department, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women, Coimbatore (Tamilnadu), India

ABSTRACT: The word "risk" is marked out as mishap that may arise in future due to some current actions. In software engineering, risk management handles variety of harms that occurs in the software due to small mistakes during the software development process. Risk management is a most influential issue which is handled by software project management (SPM). During the software life cycle many risk are related with them.

Keywords: Risk Management, Management Tools, Analysis, Monitoring, Planning.

I. INTRODUCTION

As several risks are connected with software development the cue to identify and control and manage these risks is to know about the software risk management.
There are two main types of risks are involved.
They are as follows:
-INTERNAL RISK-which can be controlled
-EXTERNAL RISK-which are beyond control

II. GOALS OF RISKMANAGEMENT

-To identify risk

-To reduce the impact and odds of risk.
-To monitor risk

III. RISK TYPES

A. Software Quality Risks
-Inadequate budget and documentation
-Lack of project standard and design
-Lack of testing and knowledge about Software.

B. Software Requirement risk
-Lack of analysis and reports for requirements

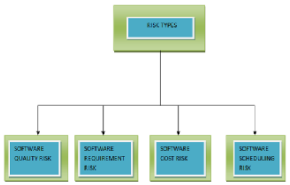


Fig. 1. Types of Risks.

-Poor definition and ambiguity of requirements.
-Invalid requirements.

C. Software Cost Risk
-Lack of testing and monitoring
-Requirement and testing changes
-Technology and Schedule change
-Malfunctioning of hardware and software tools

D. Software Scheduling Risk
-Lack of managing skills
-Inadequate knowledge and budget

IV. RISK MANAGEMENT PROCESS

The following are list of process involved in risk management:
a) Software Risk Identification
b) Software Risk Analysis
c) Software Risk Planning
d) Software Risk Monitoring

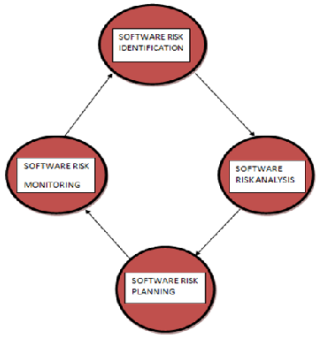


Fig. 2. Risk Management process.

a) Software Risk Identification
First step in software risk identification is to study and understand the risks which occurred in previous projects. After examining the project plans, look out for the most possible area which are exposed for risk. The outsmarted way to analyse a project plan is by converting it to a diagrammatic representation like flowchart and check all the important areas. Any resolution taken regarding the project factor must be assessed first. The most important factor is to specify the risk id, date of identification and it's description.

b) Software Risk Analysis
Software Risk Analysis is the second important step in risk management. In this step the identified risk is categorized. Later the impact, likelihood, level of categorized risk is analysed. Likelihood is defined as volume of risk which may occur due to certain technical condition like complexity of technology and low quality tools etc. Impact is defined as consequences which occur while facing risk like loss of money, customer, reputation, licence etc. There are levels of risks and they are defined high, low and medium.

c) Software Risk Planning
Software risk planning is the third step which defines the preventive methods to avoid or bring down the likelihood of various risks, impact of risks. And software risk planning monitors the process in the project for early detection of risks.

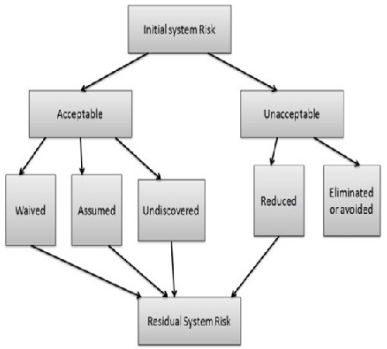


Fig. 3. Software Risk planning.

In the above Fig. 3 initially the risks are identified, and then they are categorized in to two types as acceptable and unacceptable. Acceptable risk is once again categorized in to three groups as waived (ignore), assumed and undiscovered. Unacceptable risk are either reduced or eliminated. Finally the residual or remaining risk consists of waived, assumed; undiscovered and reduced risks are monitored.

d) Software Risk Monitoring

Software risk monitoring is the fourth and final step in software risk management. The risk monitoring process is merged along with project activities. Continual checks are done upon top risks. Software risk monitoring consists of the following tasks.
-Keeping track of risk plans for any major changes.
-Preparatory measures of status report are done.
-Regular monitoring for new risks
-Evaluation risk is done. The lowest impact risk is closed.

V. RISK MANAGEMENTAPPROACH

A. Requirement analysis
First Documents are reviewed to eliminate errors. Next assessment of risk is done based on cost, schedule etc. Then failure and highly risk areas are identified and entered in risk register. Tolerance level is calculated and based on these levels requirements are prioritised.

B. Testing Process
After the analysis risk based testing is initiated so that high and medium level risks can be taken in to account for planning, implementing and monitoring. And Low level risks are kept under surveillance.
Next Quality of data is analysed and tests are planned according to levels of risks. Here various testing techniques are used in an appropriate way so that high priority risks are tested first. And different testing techniques, test data and test conditions are prepared.
The above step is followed by analysis of test plan, strategy, cases and other document generated by testing team. After performing these tasks test runs, test cases and quality check is carried out.

C. Risk Based Testing
Risk based testing is used at all the levels. There are five testing techniques to proceed risk based testing and they are as follows:
-Path Flow Testing-Ensures every path in the program are executed at least once.
-Experience Based Testing-Testing with the aid of experience gained all these years.
-Boundary Value Analysis-Testing boundaries between partitions.
-Equivalence Partitioning-Test condition is partitioned in different classes.
-Decision Tables- To test system behaviour for various inputs .Everything is represented in tabular column.
In risk based testing highly risk areas are evaluated and high risks are fully tested. Here retesting and regression testing is done to validate defect fixes and residual risks are calculated. Finally risk monitoring and control is initiated and customer feedback is obtained.

VI. RISK MANAGEMENT TOOLS AND TECHNOLOGY

A. A1 Tracker
-It is a web based tool used to record and manage the risks.
-A1 Tracker is used to build user friendly products.
-A1 Tracker is utilized to its maximum by professional users only and it is tough to learn. But it is highly opted.

-Since A1 Tracker is a web based application, it supports it assists E-mailing risk and reports to the individual who are in need.

B. Risk Management Studio

-Risk Management Studio is a flexible tool used for management of risk.
-Risk Management Studio has options like gap analysis, risk assessment and remedy.
-Risk Management Studio is an ISO 27001 certified tool.
-Risk Management Studio is easy to learn.
-It has migration assistance from excel to RM Studio and also has import, export and reporting support.

ORMStudio
Risk Management – Efficient and Effective

C. Active Risk Manager

-ARM is web based risk management tool developed by Sword Active Desk.

SwordActiveRisk

-ARM has options like recording, assessing and mitigating risks and also provides mobile application.
-ARM is used in High level establishments like Airbus, NASA etc.ARM also has menus as listed below:
-Auto Alert- Aids in broadcasting risk related updates to owner.
-Dashboard- Gives a view of various data like risks and updates in a single screen.

D. Check it

-Check it is an automated tool with audit and inspection of data options.
-Collected data is sent trough analysis, management and is reported to lower the occurrence of risk.
-Data entry in Check it tool is assisted by papers, browsers and applications on android and ios.
-Check it tool is simple and fast to learn and easy to work.
-Its customers are kelloggs and pinnacle.

checkit

E. Isometrix

-Isometrix is an application tool based upon cloud which aims mid and large levelled industries.
-This tool is used in fields like food, mining, retail, construction etc.
-It is one of the top 20 risk management tool.

ISOMETRIX
Good Business

F. Enablon

-Enablon is the most used and triumphant RM tool in current trend.
-Enablon uses two methods for tracking risks and they are top-bottom and bottom-up approach.
-It authorize user to identify, document and assess the risks.
-It has powerful internal control and management system.
-There more than thousand companies opting Enablon tool and some of them are Accenture, puma, ups etc.

enablon

G. GRC Cloud

-GRC Cloud is one of the top risk management tool developed by Resolver Systems.
-GRC Cloud has an alert system option which uses mail for triggering when risk occurs using automated technology.
-Risk, Security and incident management is handled by GRC Cloud.

RESOLVER

H. Analytica

-Analytica is an best risk management tool developed by lumina.
-It runs the model ten times faster than spreadsheet.
-Analytica is used in risk and policy analysis.
-Analytica tool generates multidimensional tables using arrays.

ANALYTICA

I. Isolocity

-Isolocity is a cloud based risk management tool which performs the process in an automated way which doesn't need any monitoring from the user.
-Hence it is a cloud based tool it can provide access to data anywhere and it is easy to learn.

ISOLOCITY

VII. FINDINGS ANDOBSERVATIONS

RISK MANAGEMENT TOOLS	ASSISTANCE PROVIDED	LEVEL AND TYPE OF TOOL	UNDERSTANDABILITY	CUSTOMERS	FINDINGS	OBSERVATIONS
A1 Tracker	Record and Manage Risk, Help Desk Staff, E-mailing Risk	TOP NOTCH	Not easy to learn, used by pro users	Dexis, Latin Asset, Beach Comber Hot Tubs	Supports Legal Contracts and Claim Contract	Manages Risks Involved in Legal and Claiming Process
Risk Management Studio	Import, Export Support and Reporting of Risk, Gap Analysis, Risk Assessment and Remedy	VERSATILE	Easy to learn and become a pro in it	Xerox, TomTom, M&G, Konica Minolta, CBOSS etc	Supports Operational Risk Management and Information Security Management	Simplifies Risk Management Through Customizable Application
Active Risk Manager	Recording, Assessing, Mitigating Risk, Auto alert, Dashboard, Quantitative and Qualitative Assessment, Mobile Application	TOP LEVEL WEB BASED APPLICATION	-	Airbus, NASA, GE Oil, Gas, US Army, Nestle, ATKINS, RWE	Supports Enterprise Risk Management	Makes Risk Management Simple, Valuable and Personal
Check it	Automated Collection of Audit and Inspection of Data. Data Entry supported by Paper, Browserand Application.	-	Easy to Use and Fast to Learn	Kellogg's, UTZ, Pinnacle etc	Supports Automatic Collection of Audit and Inspection of Data.	Minimizes the Occurrence of Risk.
Isometrix	Offers Solution like Food Safety, Occupational Health, Compliance Management, Enterprise Risk, Environmental Sustainability	TOP LEVEL CLOUD BASED APPLICATION	-	Rio Tinto, Golder Associates, METOREX, Senwes, ENVIROSERV	Supports Enterprise Risk Management	Provides Unrivalled agility and Strong Visibility
Enablon	Identification, Reporting, Documentation and Analysis	SUCCESSFUL AND FLEXIBLE APPLICATION	-	Accenture, Puma, Ups, Hydro Quebec	Supports Governance, Enterprise Risk	Offers a Comprehensive, Highly Flexible Platform

	of Risk. Provides Internal Control and Management System				Management, compliance	
GRC Cloud	Supports Risk, Security and Incident Management, Automated Alert System	TOP NOTCH CLOUD BASED APPLICATION	Easy to Learn	AIR CANADA, TELUS, ADIDAS, BASF etc	Supports Integrated Risk Management	Compliance Focused
Analytica	Multidimensional tables using arrays, Runs 10 times faster than Spreadsheet. Used in Risk and Policy Analysis.	-	Easy to learn	Environment Canada, U,S.Air Force, gsk, Nike, P&G, Unilever, Ford, Accenture,St Anford University	Supports as best solution in substitute for Complex Spreadsheet	Visual Software Environment
Isolocity	Risk Management, Opportunity, Objective, Change Managements.	CLOUD BASED APPLICATION	Easy to Learn and Use	KPT,SB tool and Machine co, Atlantic Heart Treating etc	Eliminates Labour Involved in Quality Compliance and Time Saving up to 25%	Manages Compliance

VIII. CONCLUSION

Risk Management is one of the essential areas in the development phase of a software project. It makes the software to work according to the user's requirement by making the product qualitative. In this paper we discuss about risks involved in software development, how it is managed and monitored, tools used for risk managementetc.

REFERENCES

[1]. Haneen Hijazi , Thair Khdour , Abdulsalam Alarabeyyat "A Review of Risk Management in Different Software Development Methodologies" *International Journal of Computer Applications*, Volume **45**, No.7, May 2012.

[2]. Shipra Kalra, Rachika Sharma "Quality Risk Analysis: An Approach for the Refinement of Traditional Risk Analysis" *International Journal of Engineering and Management Research*, Vol. **7**, Issue-3, May-June 2017.

[3]. Srikrishnan Sundararajan, Bhasi M., Pramod K.V. "An Empirical Study of Industry Practice in Software Development Risk Management" *International Journal of Scientific and Research Publications*, Volume **3**, Issue 6, June 2013.

[4]. Ali Rostami "Tools and Techniques in Risk Identification: A Research within SMEs in the UK Construction Industry" *Universal Journal of Management*, **4**(4): 203-210, 2016.

[5]. https://www.capterra.com/risk-management software

[6].http://www.softwaretestinghelp.com/risk-management-tools

[7].http://istqbexamcertification.com/what-is-risk-based-testinghttps://en.wikipedia.org/wiki/Risk-based_testing

[8].http://www.softwaretestingclass,com/what-is-risk- based-testing-in-software-testing

[9].https://www.guru99.com/risk-based-testing.html

[10].http://www.test-institute.org/What_Is_Software_Risk_And_Software_Risk_Management.php

[11].http://www.softwaretestingclass.com/what-is-risk-based-testing-in-software-testing.

[12].http://www.professionalqa.com/experience-based-testing.

[13].https://www.guru99.com/equivalence-partitioning-boundary-value-analysis.html

[14].https://www.guru99.com/software-testing- techniques-1.html

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 58-62(2018)

International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Speech Recognition for Speakers with Dysarthria using TORGO Database

Usha. M[1] and Dr. L. Sankari[2]

[1]Assistant Professor & Head, Associate Professor,
Department of Computer Applications, KG College of Arts & Science, Coimbatore (Tamilnadu), India
[2]Associate Professor, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women, Coimbatore (Tamilnadu), India

ABSTRACT: Automatic speech recognition(ASR) is being used in a range of assistive contexts, comprising home computer systems, mobile telephones, and various public and private telephony facilities. Dysarthria is a motor speech disorder in which the person lacks the control over articulators used for speech production. Speech accuracy is the outcome of well-timed and coordinated activities of the articulators and other related neuro muscular feature. Speech of the dysarthric disordered people is assessed using structured sparse feature selection and prediction. In this proposed work, Speech utterance is converted into a phone sequence and histograms of the pronunciation mappings are done by using Mel-frequency cepstral coefficients. Structured sparse feature selection is done using Hidden Markov Models. Prediction is done using Inverse Mel-frequency cepstral coefficients. The TORGO database is mostly a resource for emerging ASR models more suitable to the requirements of people with atypical speech production, although it is similarly useful to the more general ASR community. TORGO Database is used here to identify the dysarthric disabled person.

Keywords : Dysarthria, Sparse feature selection, MFCC, Hidden Markov Models.

I. INTRODUCTION

Automatic speech recognition is being used in a range of assistive contexts, comprising home computer systems, mobile telephones, and various public and private telephony facilities. Despite their developing presence, commercial speech recognition technologies are quiet not easily hired by individuals who have speech or communication disorders. While speech disorders in older adults are common, there has been reasonably little research on automatic speech recognition performance with older adults. However, research findings advise that the speech characteristics of the older adult may, in some ways, be related to dysarthric speech. Dysarthria, a common neuro-motor speech disorder, is particularly valuable for exploring automatic speech recognition performance limitations because of its widespread range of speech expression.

Dysarthria is a set of inherited and traumatic neuromotor disorders that harm the physical production of speech. These impairments decrease or remove normal control of the primary vocal articulators but do not disturb the regular comprehension or production of meaningful, syntactically correct language. Congenital causes of dysarthric speech are often initiated by some sort of asphyxiation of the brain, impeding normal development in the speech–motor areas [1].

A feature representation method based on alignment with the decoded (spoken) phone sequence resulting from an ASR system and canonical phone sequence from a standard pronunciation dictionary. To this end, a weighted finite state transducer (WFST) [2] is employed as an alignment tool to capture phone-level mappings including match, substitution, and deletion. A WFST is a very powerful and flexible framework in mapping input and output symbols, and therefore, it has been utilized in various speech and language processing applications such as machine translation, pronunciation modeling, and effi- cient speech recognition [3].

Feature selection and intelligibility scoring methods based on structured sparse linear models. Sparse linear models seek to predict an output by linearly combining a small subset of the features.

The TORGO database is mostly a resource for emerging ASR models more suitable to the desires of people with atypical speech production, although it is similarly useful to the more general ASR community. A primary motive for collecting detailed physiological information is to be gifted to explicitly learn 'hidden' articulatory parameters automatically in computer speech models via statistical pattern recognition. This database is also suitable in the clinical domain, especially by linguists and pathologists involved in studying atypical speech production.

II. BACKGROUND STUDY

A. Types of Automatic Speech Recognition Systems

There are basically three categories of ASR systems differentiated by the degree of user training required prior to use: (1) speaker dependent, (2) speaker

independent, and (3) speaker adaptable ASR. Speaker dependent ASR requires speaker training or enrollment prior to use and the primary user trains the speech recognizer with samples of his or her own speech. These systems typically work well only for the person who trains it. Speaker independent ASR does not require speaker training prior to use. The speech recognizer is pre-trained during system development with speech samples from a collection of speakers. Many different speakers will be able to use this same ASR application with relatively good accuracy if their speech falls within the range of the collected sample; but ASR accuracy will generally be lower than achieved with a speaker dependent ASR system. Speaker adaptable ASR is similar to speaker independent ASR in that no initial speaker training is required prior to use. However, unlike speaker independent ASR systems, as the speaker adaptable ASR system is being used, the recognizer gradually adapts to the speech of the user.

B. Dysarthria and Older Adult Speech

Dysarthria, a neuro-motor speech disorder, may arise secondary to diseases such as Parkinson's, multiple sclerosis, and amyotrophic lateral sclerosis; disorders such as right hemisphere syndrome or dementia; or following traumatic brain injury or stroke. Several types of dysarthria exist, each of which has different expressed speech characteristics. Typically, dysarthria is classified according to the site of lesion and degree of neurological damage; however, in the literature reviewed, dysarthria is loosely classified based on the degree of disorder severity as measured by speech intelligibility and articulation. For example, mild, moderate and severe classifications were used as opposed to site of lesion [4].

Raghavendra *et al.* (2001) examined ASR accuracy results obtained from four dysarthric speakers (mild, moderate, severe and profoundly severe) and one control speaker, using a speaker dependent, discrete word, whole-word pattern matching ASR system (Infovox RA) and a speaker adaptable, discrete word, phoneme-based ASR system (Swedish Dragon Dictate). Degree of dysarthria was determined using the Swedish Dysarthria Test. For both the Infovox RA and the Swedish Dragon Dictate systems, the average accuracy ratings over three sessions of use, was highest for the control and mildly dysarthric speakers, followed by the moderately dysarthric, then severely dysarthric, and finally the profoundly severe dysarthric speaker. Generally, all speakers achieved higher accuracy ratings using Swedish Dragon Dictate (74%- 97%) over Infovox RA (28%-97) [5].

The TORGO database of dysarthric articulation consists of aligned acoustics and measured 3D articulatory features from speakers with either cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS), which are two of the most prevalent causes of speech disability, and matchd controls. This database, called TORGO, is the result of a collaboration between the departments of Speech-Language Pathology at the University of Toronto and the Holland-Bloorview Kids Rehab hospital in Toronto.

Both CP and ALS result in *dysarthria*, which is caused by disruptions in the neuro-motor interface. These disruptions distort motor commands to the vocal articulators, resulting in atypical and relatively unintelligible speech in most cases. This unintelligibility can significantly diminish the use of traditional automatic speech recognition (ASR) software. The inability of modern ASR to effectively understand dysarthric speech is a major problem, since the more general physical disabilities often associated with the condition can make other forms of computer input, such as keyboards or touch screens, especially difficult.

III. METHODOLOGY

A. Feature Selection

Here feature selection is done using structured sparse linear feature selection using Hidden Markov Model (HMM).

A new feature selection strategy to choose a proper feature set, which predicts accurately subjective intelligibility scores, from a number of features. The proposed selection criterion is in the form of penalty-based objective function with its associated weighting parameter for the purpose of selecting proper features which not only produce a small prediction error but also keep their mutual dependency low. In the method, we find the best feature set by iteratively selecting one feature satisfying the feature selection criterion. To this end, the prediction error of m-th feature at t-th round is calculated based on the 0-1 step loss function defined by

$$\varepsilon_{t,n} = 2\left[\{1+\exp(-\alpha\cdot\gamma_{t,n})\}^{-1}-1\right],\ \alpha>0 \qquad (1)$$

where

$$\gamma_{t,n} = \left[\sum_{i}\left(y_i-\hat{y}_{n,n}\right)^2/N\right]^{1/2}. \qquad (2)$$

N is the number of training speakers, yn is the intelligibility score of n-thdysarthric speaker, $\hat{y}_{m,n}$ is the predicted score of n-thdysarthric speaker based on a multiple linear regression model using both already selected features at the previous round and m-th feature to be selected at current t-th round [6].

An Hidden Markov Model(HMM) based classifier system usually incorporates one individual HMM for each class where each HMM is built up from a set of states. For all these states output distributions, and for all pairs of states transition probabilities are defined. The input to an HMM classifier is a sequence of feature vectors x1,...,xn. In the recognition, or decoding, phase the input sequence x1,...,xn is mapped to a sequence of states si1,...,sin and for each such mapping a likelihood value is defined by the HMM. The optimal mapping, which maximizes the likelihood, is usually found by means of the Viterbi algorithm. This optimal mapping is equivalent to an optimal path through a graph that is defined by the product of the sequence of input vectors and the states of the HMM.The likelihood of the optimal path is the score of the sequence of feature vectors for the considered HMM [7].

B. Prediction

Prediction is done using Mel Frequency Cepstral Coefficient(MFCC). A correlation analysis is performed by first defining an acoustic speech feature vector, fi , which comprises the fundamental frequency, f0, and the frequencies of the first four formants, F1, F2, F3, and F4, at time frame i, fi = f0,F1,F2,F3,F4.

For voiced speech, multiple correlations are measured between each element of the acoustic feature vector and the MFCC vector. For unvoiced speech, multiple correlations are only calculated between formant frequencies and MFCC vectors. The multiple correlation between each acoustic speech feature and the MFCC vector is measured using multiple linear regression.8 A linear model is computed to describe the relationship between MFCC vectors independent variables and acoustic speech feature vectors dependent variables. Each acoustic feature at frame i, fi j, is represented in terms of the ith MFCC vector,xi=xi…..1,xi……2,...,xi, using a set of M + 1 regression coefficients, 0,j ,...,m,j ,...,M,j, which are specific to the jth acoustic feature

$$f_i(j) = \beta_{0,j} + \beta_{1,j}x_i(1) + \beta_{2,j}x_i(2) + \cdots + \beta_{M,j}x_i(M) + \varepsilon$$

$$0 \le i \le I-1, \quad 1 \le j \le 4 \quad \text{for unvoiced speech,}$$

$$0 \le i \le I-1, \quad 0 \le j \le 4 \quad \text{for voiced speech,}$$

where is an error term, I is the number of MFCC vectors, and M is the dimensionality of the MFCC vector.[8]

C. TORGO Database

All subjects read English manuscript from a 19-inch LCD screen. One subject experienced some visual tiredness near the end of one session, and therefore repeated a small section of verbal stimuli spoken by an experimenter. No perceptible effect of this approach was measured. The stimuli were presented to the participants in randomized order from within fixed-sized collections of stimuli in order to keep left from priming or dependency effects. Dividing the stimuli into collections in this manner guaranteed overlie between subjects who speak at vastly different rates. Stimuli are classified into the following categories:

Non-words. These are used to organize for the baseline abilities of the dysarthric speakers, especially to gauge their articulatory control in the presence of plosives and prosody. Speakers are asked to attain the following:

- Repetitions of /iy-p-ah/, /ah-p-iy/, and /p-ah-t-ah-k-ah/. These progression allow us to monitor phonetic contrasts around plosive consonants in the occasion of high and low vowels.
- High-pitch and low-pitch vowels. This allows us to determine the use of prosody in assistive communication [8].

Short words. These are helpful for studying speech acoustics without the need for word boundary detection. This grouping includes the following:

- Repetitions of the English digits, 'yes', 'no', 'up', 'down', 'left', 'right', 'forward', 'back', 'select', 'menu', and the international radio alphabet (e.g., 'alpha', 'bravo', 'charlie'). These words are helpful for hypothetical command software for accessibility.
- 50 words from the the word intelligibility section of the Frenchay Dysarthria Assessment.
- 360 words from the word intelligibility section of the Yorkston-Beukelman Assessment of Intelligibility of Dysarthric Speech (Yorkston and Beukelman, 1981).
- The 10 most universal words in the British National Corpus.
- All phonetically contrasting pairs of words from Kent et al. (1989). These are grouped into 18 articulation-relevant categories that affect intelligibility, including glottal/null, voiced/voiceless, alveolar/palatal fricatives and stops/nasals.

TORGO Directory and file structure. All data is ordered by speaker and by the session in which each speaker recorded data.

Speaker data. Each speaker is assigned a code and precise their own directory. Female speakers have a code that begins with 'F' and male speakers have a code that begins with 'M'. If the speaker is a member of the control group (i.e., they do not have a form of dysarthria), then the letter 'C' follows the gender code. The last two digits simply indicate the order in which that subject was recruited. For example, speaker 'FC02' is the second female speaker without dysarthria recruited.

All subjects read English manuscript from a 19-inch LCD screen. One subject experienced some visual tiredness near the end of one session, and therefore repeated a small section of verbal stimuli spoken by an experimenter.
Each speaker's directory contains 'Session' directories, which encapsulate data recorded in the respective visit, and occasionally a 'Notes' directory which can include Frenchay assessments, notes about sessions (e.g., sensor errors), and other relevant notes.
Each 'Session' directory can contain the following content:

alignment.txt This is a text file containing the sample offsets among audio files recorded simultaneously by the array microphone and the head-worn microphone. The first line is a space-separated pair of directories representing that indicated offsets refer to files in the second directory relative to those in the first. All subsequent lines in alignment.txt indicate the common filename and the sample offset, separated by a space.

amps/ These directories contain raw *.amp and *.ini files produced by the AG500 articulograph.

phn_*/ These directories contain phonemic transcriptions of audio data. Each file is plain text with a *.PHN file extensions and a filename referring to the utterance number. These files were generated using the free **Wavesurfer** tool according to the TIMIT phone set, with phonemes marked *cl referring to closures before plosives. Files in 'phn_arrayMic' are aligned temporally with acoustics recorded by the array microphone and files in 'phn_headMic' are aligned temporally with acoustics recorded by the head-worn microphone.

pos/ These directories contains the head-corrected positions, velocities, and orientations of sensor coils for each utterance, as generated by the AG500 articulograph. These files can be read by the 'loaddata.m' function in the included 'tapadm' toolkit (see below) and contain the primary articulatory data of interest. Except where noted, the channels in these data refer to the following positions in the vocal tract:

1. Tongue back (TB)
2. Tongue middle (TM)
3. Tongue tip (TT)
4. Forehead
5. Bridge of the nose (BN)
6. Upper lip (UL)
7. Lower lip (LL)
8. Lower incisor (LI)
9. Left lip
10. Right lip
11. Left ear
12. Right ear

prompts/ These directories contain orthographic transcriptions. Each filename refers to the utterance number. Prompts marked 'xxx' indicate spurious noise or otherwise generally unusable content. Prompts indicating a *.jpg file refers to images in the Webber Photo Cards: Story Starters collection.

rawpos/ These directories are equivalent to the pos/ directories except that their articulographic content is not head-normalized to a constant upright position.

wav_*/ These directories contain the acoustics. Each file is a RIFF (little-endian) WAVE audio file (Microsoft PCM, 16 bit, mono 16000 Hz). Filenames refer to the utterance number. Files in 'wav_arrayMic' are recorded by the array microphone and files in 'wav_headMic' are recorded by the head-worn microphone.

IV. RESULTS AND DISCUSSION

Ten different speech samples are examined.The result values of R and RMSE are given in the following table:

Table 1: speech data.

Speech Data	RMSE	R
Speech Data 1	17.001	1.1314
Speech Data 2	31.72	2.8156
Speech Data 3	18.5	0.2212
Speech Data 4	12.45	0.121
Speech Data 5	35.67	3.1023
Speech Data 6	21	1.23
Speech Data 7	15.2	0.78
Speech Data 8	26.72	2.3156
Speech Data 9	20.45	2.09
Speech Data 10	16.45	1.53

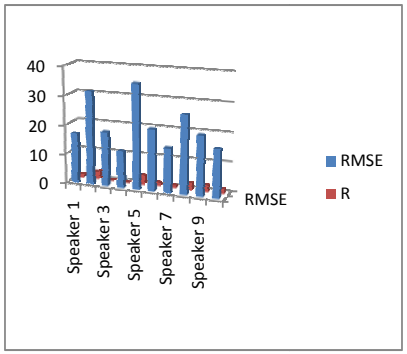


Fig. 1. Pictorical representation of the data.

The Chart shows that the following classification of the input sample:

Classification	Speech Data
Normal	Speech Data 1
	Speech Data 6
	Speech Data 10
Controlled	Speech Data 3
	Speech Data 4
	Speech Data 7
	Speech Data 2
Dysarthric	Speech Data 5
	Speech Data 8
	Speech Data 9

The Data of the speaker 1,6,10 are classified as Normal Speaker depnds on The R and RMSE values. The speaker 3,4,7 are classified as controlled speakers and the speaker 5,8,9 are classified as Dysarthric speakers.

V. CONCLUSION

In this paper. Assessment of the dysarthric disabled people were made easy because of the structured sparse feature selection and prediction. Structured sparse feature selection is done using Hidden Markov Models. Prediction is done using Inverse Mel-frequency cepstral coefficients. Torgo database helps in identifying the Male and femailedysarthric disordered people effectively. In order to reduce the noise, adaptive noise cancelling approach can be applied in future.

REFERENCES

[1]. Young V, Mihailidis A, "Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: a literature review." 2010 Summer; **22**(2): 99-112.
[2]. Myungjong Kim, YounggwanKim,"Automatic Intelligibility Assessment of Dysarthric Speech Using Phonologically-Structured Sparse Linear Model"in IEEE/ACM Transactions on Audio, Speech, and Language Processing, **23**(4): 694-704 · April 2015.
[3] M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," Comput. Speech Lang., vol. **16**, no. 1, pp. 69–88, 2002.
[4]. Victoria Young, Alex Mihailidis, "Difficulties in Automatic Speech Recognition of Dysarthric Speakers and the Implications for speech based applications used by the elderly: a Literature review"
[5]. Raghavendra, P., Rosengren, E., & Hunnicutt, S. (2001),"An investigation of different degrees of dysarthric speech as input to speaker-adaptive and speaker-dependent recognition systems", *Augmentative and Alternative Communication*, **17** (4), 265-275.
[6]. M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," Comput. Speech Lang., vol. **16**, no. 1, pp. 69–88, 2002.
[7]. M. Hasegawa-Johnson, J. Gunderson, A. Perlman, T. Huang**,** "HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria"**,** in: Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, pp. 1060–1063.
[8]. Kinfe Tadesse Mengistu and Frank Rudzicz"Comparing Humans and Automatic Speech Recognition Systems in Recognizing Dysarthric Speech", University of Toronto, Department of Computer Science.

International Journal of Theoretical & Applied Sciences, Special Issue 10(1a): 63-67(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Review on Feature Selection Approaches for Heart Disease Classification

C. Usha Nandhini[1] and Dr. P.R. Tamilselvi[2]
[1]Part Time Research Scholar in Periyar University, Assistant Professor of Computer Applications, Vellalar College for Women (Autonomous), Erode (Tamilnadu), India,
[2] Assistant Professor of Computer Science, Govt. Arts and Science College, Komarapalayam (Tamilnadu), India

ABSTRACT: Heart disease describes a range of conditions that affect heart. It refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain or stroke. Diagnosis of heart disease is expensive as many tests are required to predict the disease. By using data mining techniques we can reduce the cost of diagnosis by avoiding many tests by selection of those attributes which are really important for prediction of disease. As medical information is with multiple attributes, medical data mining differs from other one. To achieve the classification accuracy, we need to remove the redundant and the irrelevant data present. There are a number of feature selection methods available in literature due to the availability of data with many variables leading to data with very high dimension. Feature selection methods provides us a way of reducing computation time, improving prediction performance, and a better understanding of the data in machine learning algorithms. This work presents a comprehensive review for the feature selection approaches for the prediction of heart disease.

Keywords: Heart Disease, Attribute selection, Classification.

I. INTRODUCTION

Data sets for analysis may contain hundreds of attributes, many of which may be irrelevant to the mining task or redundant. Although it may be possible for a domain expert to pick out some of the useful attributes, this can be a difficult and time-consuming task, especially when the behavior of the data is not well known. Leaving out relevant attributes or keeping irrelevant attributes may be detrimental, causing confusion for the mining algorithm employed. This can result in discovered patterns of poor quality. In addition, the added volume of irrelevant or redundant attributes can slow down the mining process. Feature subset selection reduces the data set size by removing irrelevant or redundant attributes. The goal of feature subset selection is to find a minimum set of features such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on a reduced set of attributes has an additional benefit. It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand [1].

A. Greedy Method

For n attributes, there are 2^n possible subsets. An exhaustive search for the optimal subset of attributes can be prohibitively expensive, especially as n and the number of data classes increase. Therefore, heuristic methods that explore a reduced search space are commonly used for feature subset selection. These methods are typically greedy in that, while searching through attribute space, they always make what looks to be the best choice at the time. Their strategy is to make a locally optimal choice in the hope that this will lead to a globally optimal solution. Such greedy methods are effective in practice and may come close to estimating an optimal solution. The best and worst features are typically determined using tests of statistical significance, which assume that the features are independent of one another. Many other feature selection measures can be used, such as the information gain measures used in building decision trees for classification [1].
Basic heuristic methods of feature subset selection include the following techniques, some of which are listed below[1]:

1. Stepwise forward selection: The procedure starts with an empty set of features as the reduced set. The best of the original features is determined and added to the reduced set. At each subsequent iteration or step, the best of the remaining original features is added to the set [1].
2. Stepwise backward elimination: The procedure starts with the full set of features. At each step, it removes the worst attribute remaining in the set [1].
3. Combination of forward selection and backward elimination: The stepwise forward selection and backward elimination methods can be combined so that, at each step, the procedure selects the best feature and removes the worst from among the remaining features [1].
4. Decision tree induction: Decision tree algorithms, such as ID3, C4.5 and CART, were originally intended for classification. Decision tree induction constructs a flowchart like structure where each internal node denotes a test on an attribute, each branch corresponds to an outcome of the test, and each external node denotes a class prediction. At each node, the algorithm chooses the "best" attribute to partition the data into individual classes [1].
   When the decision tree induction is used for feature subset selection, a tree is constructed from the given data. All features that do not appear in the tree are assumed to be irrelevant. The set of features appearing in the tree form the reduced subset of features. The stopping criteria for the methods may vary. The procedure may employ a threshold on the measure used to determine when to stop the feature selection process [1].

II. LITERATURE REVIEW

Authors in paper [2] investigated the significance of feature selection methods for improving the performance of classification methods. The experimentation is conducted on dataset of health care domain. The statlog heart disease data is collected from UCI Machine Learning Repository. In this work, the hybrid attribute selection method combining CFS and Filter Subset Evaluation techniques are applied to select the relevant features. They also propose a new feature selection method algorithm which is another hybrid method combining CFS and Bayes Theorem. It is found that the CFS and FILTER SUBSET EVALUATION reduces more number of irrelevant and redundant attributes thereby increases the performance of classifiers. It selects 6 features namely 3-chest, 8-maximum_heart_rate_achieved , 9-exercise_induced_angina,12- no_major_vessels, 13-thal. The new proposed feature selection namely CFS and Bayes Theorem selects only 3 features(3-chest, 12- number_of_major_vessels, 13- thal) that gives better accuracy for NB and KNN classifier. They conclude that CFS and BAYES THEOREM based feature selector is best suitable for heart disease data prediction.

Authors in paper [3] present a new method for the diagnosis of Coronary Artery Diseases (CAD) founded on Genetic Algorithm (GA) wrapped Bayes Naïve (BN) based Feature Selection. Basically, Coronary Artery Diseases (CAD) dataset contains two classes defined with 13 features. The automatic process to generate the subset of features is an advantage of the proposed algorithm. The proposed algorithm for CAD patients contains two steps such as: (1) generation of a subset of features and (2) and the evaluation of the system using Bayes Naïve Machine Learning technique. The experimental results demonstrate the strength of the proposed GA wrapped BN algorithm for selecting the most relevant features for efficient diagnosis of CAD diseases. It selects 7 features such as 1-cp, 3-sex, 7-restescg, 10-oldpeack, 11-slope, 12-ca, 13-thal. In order to evaluate the efficacy of our proposed GN wrapped BN technique, they compare their methodology with another two FS wrapped methodologies. The two methodologies are based on BN algorithm. The first one uses the Best First Search (BFS) as a generation technique. BFS is a search algorithm that explores a graph by expanding the most promising node with the best score which will be evaluated using the wrapped BN [29].It selects 6 features such as 5-chol, 6-fbs, 8-thalach, 9-exang, 12-ca, 13-thal. In the second methodology they generate the subset of features using the Sequential Floating Forward Search (SFFS). This technique is derived from the sequential forward generation techniques. The principle of such techniques is to add one or more attributes progressively. However, as they do not explore all possible subsets of attributes and

cannot backtrack during the search, so they are suboptimal. SFFS after each step Forward, it applies Backward steps while the subset corresponding improves the efficacy of wrapped BN. It selects 6 features such as 1-cp, 7-restescg, 8-thalach, 10-oldpeack, 12-ca, 13-thal. Generally automated disease diagnosis problems need a reduction of Features space step to achieve high accuracy performance. The final set of attribute contains the most relevant feature model that increases the accuracy. Consequently, the proposed algorithm is applied to CAD disease. Thus, the asset of the algorithm is then compared with the use of Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and C4.5 decision tree Algorithm. The results of classification accuracy for those algorithms are respectively 83.5%, 83.16% and 80.85%. Consequently, the GA wrapped BN Algorithm is correspondingly compared with other FS algorithms.

Author of this paper [4] analyzed the approach of feature selection for classification and also presented a novel approach for the feature selection by using association and correlation mechanism. The aim of this paper is to select the correlated features or attributes of medical dataset so that patient need not to go for many tests and in future it is used for preparing the clinical decision support system which is helpful for decision making of disease prediction in a cheaper way. Other approach is mentioned in this paper is after removal of some attributes, accuracy of classifier is also improved which support statement of disease prediction in cheaper way by avoiding all unwanted tests for disease prediction. By using association rules and correlation attributes features can be selected. As medical field contains large number of attributes and information so dimensionality reduction is must now. The accuracy of classifiers after removal of attributes is discussed in this paper. The features selected during this approach for Heart disease datasets are applied into Neural network for classification For NN attribute subset as R={2-sex,5-chol,6-fbs,8-thalach,9-exang,11-slope,12-ca,13-thal} are important with these attributes accuracy is more as 84.4.9%. , using Decision tree the selected features are 6 as R= {1-age, 3-cp, 9-exang, 11-slope, 12-ca, 13-thal} with accuracy 84.16% and using SVM the accuracy of classifier is more 87.46% for features identified as R= {2-sex, 3-cp, 6-fbs, 9-exang, 10-oldpeak, 11-slope, 12-ca, 13-thal} for same dataset. By removing irrelevant attributes the performance of classification can be improved and also cost of classification may get reduced. In this work the analysis of two different approaches for feature selection is done specially for medical datasets. They conclude that feature selection really helpful for dimensionality reduction and also for building cost effective model for disease prediction.

Authors in paper [5] build a classifier that will predict the presence or absence of a disease by learning from the minimal set of attributes that has been extracted from the clinical dataset. In this work rough set indiscernibility relation method with back propagation neural network (RS-BPNN) is used. This work has two stages. The first stage is handling of missing values to obtain a smooth data set and selection of appropriate attributes from the clinical dataset by indiscernibility relation method. The second stage is classification using back propagation neural network on the selected reducts of the dataset. The classifier has been tested with hepatitis, Wisconsin breast cancer, and Statlog heart disease datasets obtained from the University of California at Irvine (UCI) machine learning repository. The accuracy obtained from the proposed method is 97.3%, 98.6%, and 90.4% for hepatitis, breast cancer, and heart disease, respectively. The proposed system provides an effective classification model for clinical datasets. The rough set theory is used to produce minimal subset of attributes to represent the whole features of the dataset. From roughset theory the proposed method used indiscernibility relation method to select the reducts. From heart disease dataset the proposed method produced 4473 reducts with attribute of 3 to 12. Nine reducts and all attributes from each of the datasets are used for comparison of classification result. An artificial neural network with back propagation learning algorithm is used for classification.The best performance result for heart disease is obtained from Reduct-R6 with 6 attributes (Chp, Sch, Ecg, Opk, vessel,andThal). It achieves 90.4% accuracy. The proposed work achieves superior performance when compared to recent and conventional works. It is not always true that the reduct with less number of attributes gives highest classification accuracy. In this study heart disease dataset have got their best accuracy performance with the reduct that consists of 6 attributes. In future, the use of hybrid methods of roughset theory with optimization techniques like particle swarm optimization (PSO), ant colony optimization (ACO), genetic algorithm (GA), and bacterial foraging (BF) can be experimented with many more datasets.

The authors [6] introduced a new approach for prediction problem with the objective of attaining maximum classification accuracy with smallest number of features selected. Cardiac diseases are very common and one of the main reasons of death. Hence the main objective of this paper is to predict the possibility of heart disease at its early stages with less number of attributes. Their approach integrates anthropometric data and physiological data of cardiac diseases by proposing novel feature selection method for prediction of heart diseases. The dataset used in this work is collected from Cleveland Heart disease database. The novel feature selection extracted five major features such as Trestbps, Restecg, Slope, CA, thal and age. This selection of minimum features is used to predict the possibility of heart diseases at its early stages and generate a alarm for physicians and patient as well. They ran experiments in all data mining algorithms and proved that neural network predicts 93% of accuracy and Sequential Minimal Optimization (SMO) predicts 89% of accuracy. The results show the proposed approach leads to a superior feature selection process in terms of sinking the number of variable required and increased in classification accuracy for better prediction. This decision support system can use for providing better healthcare services to heart patient. Thus the early diagnosis of heart disease detection may decrease the chances of death in cardiac.

The authors [7] proposed a novel ReliefF and Rough Set (RFRS) based classification system for heart disease diagnosis. The proposed system contains two subsystems: the RFRS feature selection system and a classification systemwith an ensemble classifier.Thefirst system includes three stages: (i) data discretization, (ii) feature extraction using the ReliefF algorithm, and (iii) feature reduction using the heuristic RoughSet reduction algorithm that we developed. In the second system, an ensemble classifier is proposed based on the C4.5 classifier. The Statlog (Heart) dataset, obtained from the UCI database, was used for experiments. The experimental results show that the reduct$R2$ has seven attributes ($C1$- Age,3-Chest pain type, $C7$-Resting electrocardiographic results, $C8$-Maximum heart rate achieved, $C11$-Slope of the peak exercise ST segment, $C12$Number of major vessels (0–3) colored by fluoroscopy, $C13$-Thal) achieves the highest classification accuracy (92.59%) using an ensemble classifier with the C4.5decision tree as the weak learner. The results also show thatthe RFRS method has superior performance compared tothree common classifiers in terms of ACC, sensitivity, andspecificity. In addition, the performance of the proposedsystem is superior to that of existingmethods in the literature.Based on empirical analysis, the results indicate that theproposed classification system can be used as a promisingalternative tool in medical decision making for heart disease diagnosis. However, the proposed method also has some weaknesses.The number of the nearest neighbors (k) and theweight threshold (θ) are not stable in the ReliefF algorithm. One solution to this problem is to compute estimatesfor all possible numbers and take the highest estimate of each feature as the final result. They need to perform moreexperiments to find the optimal parameter values for the ReliefF algorithm in the future.

III. CONCLUSION

This survey provides the brief description of feature selection techniques for classification of heart disease. The classification accuracy depends on the exact metrics which indicates the number of features has been utilized. Analysis presented by different researcher's shows that different feature selection methods and classifiers are defined which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that using different techniques and taking different number of attributes we get different accuracies for predicting heart diseases.

REFERENCES

[1]. Han, J., Kamber, M. (2006). "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.

[2]. John Peter. T, K. Somasundaram, (2012). "Study and Development of Novel Feature Selection Framework for Heart Disease Prediction", *International Journal of Scientific and Research Publications*, Volume **2**, Issue 10, October 2012, ISSN 2250-3153

[3]. Sidahmed Mokeddem, Baghdad Atmani and Mostéfa Mokaddem, (2013). "Supervised feature selection for Diagnosis of Coronary Artery Disease Based on Genetic Algorithm", Sundarapandian et al. (Eds): CSE, CICS, DBDM, AIFL, SCOM – 2013, pp. 41–51, 2013. © CS & IT-CSCP 2013.

[4]. Prof. K. Rajeswari, Dr. V. Vaithiyanathan and Shailaja V. Pede, (2013). "Feature Selection for Classification in Medical Data Mining", *International Journal of Emerging*

Trends & Technology in Computer Science, Volume **2**, Issue 2, March – April 2013 ISSN 2278-6856.
[5]. Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, and Kannan Arputharaj, (2015). "Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network", Hindawi Publishing Corporation, *Computational and Mathematical Methods in Medicine*, Volume **2015**, Article ID 460189, 13 pages.
[6]. Suganya. R, Rajaram. S, Sheik Abdullach .A and Rajendran. V, (2016). "A Novel Feature Selection Method for Predicting Heart Disease with Data Mining Techniques", *Asian Journal of Information Technology* **15**(8) , pp: 1314-1321, ISSN: 1682-3915.
[7]. Xiao Liu, Xiaoli Wang, Qiang Su, Mo Zhang, Yanhong Zhu, Qiugen Wang, and Qian Wang, (2017). "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method", Hindawi Publishing Corporation, Computational and Mathematical *Methods in Medicine*, Volume 2017, Article ID 8272091, 11 pages.
[8]. R. Subha, (2015). "Study on Cardiovascular Disease Classification using Machine Learning Approaches", *IJECS*, Volume **04**, Issue 12, December 2015, Page No. 15177-15182.
[9]. Samir Roy, Udit Chakraborty, (2013). "Introduction to Soft Computing", Pearson India Education Services Pvt. Ltd., 2013.
[10]. Gupta. G.K., (2008). "Introduction to Data Mining with Case Studies", Prentice Hall of India Private Limited, New Delhi, 2008.
[11]. Nancy. P, Sudha. V, Akiladevi. R, (2017). "Analysis of feature selection and classification algorithms on Hepatitis Data", *IJARCET*, Volume **6**, Issue 1, January 2017. Pages 19-23.
[12]. Rajeswari. K, Dr. V. Vaithiyanathan, Dr. T.R. Neelakantan, (2012). "Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks", Elsevier, *International Symposium of Robotics and Intelligent Sensors* 2012, Pages 1818-1823.
[13]. Mustafa Serter Uzer, Nihat Yilmaz and Onur Inan, (2013). "Feature Selection Method Based on ABC Algorithm and SVM for Medical Datasets Classification, Research Article, Hindawi Publishing Corp., *The Scientific World Journal*, Vol. 2013.

International Journal of Theoretical & Applied Sciences, Special Issue 10(1a): 68-72(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Optimization of Support Vector Machine Parameters using Grid Search Method

Premasundari M[1] and Dr. C.Yamini [2]
[1]Research Scholar, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women Coimbatore (Tamilnadu), India.
[2]Associate Professor, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women Coimbatore (Tamilnadu), India.

ABSTRACT: Classification technique is one of the most important machine learning techniques to separate the information into different groups or categories. There are many classification algorithms such as Decision tree, Support vector machine, Naïve Bayes, and Artificial Neural Network. Support Vector Machine (SVM) is a learning algorithm which can be applied for both categorization and regression challenges. In this composition, support vector machine is applied for classification of the Iris dataset in R tool and the parameters of SVM are optimized to get the optimal result. The optimized results are more serious than the initial consequences of the proposed classifier.

Keywords: Classification, Support Vector Machine, R tool, Regression

I. INTRODUCTION

Machine learning is defined by Arthur Samuel in 1959. Machine learning is the arena of computer science that developed from the work of computational learning theory in artificial intelligence. It explores the subject area and building of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model of a sample training set of input data to create data-driven predictions or decisions expressed as outputs.

A. Types of Machine Learning

Supervised learning, Unsupervised learning and Semi-supervised learning are the dissimilar cases of machine learning. A supervised learning algorithm analyses the training data and produces an inferred function. Unsupervised learning is the task of inferring a function to describe concealed structure from unlabeled data [1]. Semi-supervised learning is a course of study of supervised learning methods and techniques that also gain use of unlabeled data for training. Semi-supervised learning lies between unsupervised learning and supervised learning

B. Applications of Machine Learning

Machine Learning can be applied in the following fields:

- Bioinformatics
- Computer vision
- Data Mining
- Natural Language Processing
- Pattern Recognition
- Speech Recognition
- Recommendation system

II. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are supervised learning methods used for classification and regression tasks that originated from statistical learning theory [2]. Support Vector Machines were extremely popular around the time and a high-performing algorithm with slight tuning. It is an exciting algorithm and the concepts are quite simple. A hyperplane is a line that splits the input variable space. In SVM, a hyperplane is selected to separate the data points in the input variable space by its class, either class 0 or class 1. The Separation of the input points can be visualized in n-dimensional space. SVM is normally intended for binary classification problems and further improved for regression and multi-class classification.

A. Maximal-Margin classifier

The Maximal-Margin Classifier is an imaginary classifier that describes the working of SVM [3]. The distance between the line and the closest data points is referred to as the margin. The optimal line that can separate the two classes is called the Maximal-Margin hyperplane. The margin is calculated as the perpendicular distance from the line to only the closest points. But these points which are relevant in defining the line and constructing the classifier are called support vectors. The support vectors define the hyperplane. The hyperplane is learned from training data using an optimization procedure that maximizes the margin.

B. Soft-Margin Classifier

Real data are chaotic and cannot be separated perfectly with a hyperplane [3]. The constraint of

maximizing the margin of the line that divides the categories must be loosened. This is frequently called the soft margin classifier. This change allows some points in the training data to violate the separating line.

C. Kernels

Normally, the SVM algorithm is executed using a kernel. The learning of the hyperplane in linear SVM is performed by transforming the problem into some linear algebra. The kernel defines the similarity or a distance measure between new data and the support vectors. The similarity measure is used for linear kernel because the distance is a linear combination of the inputs. Other kernels such as a Polynomial and a Radial Kernel can be used to transform the input space into higher dimensions. This is called the Kernel Trick. More complex kernels can be used to separate the groups and thus provides more accurate classifiers. The equations of common kernel functions are shown in Table 1 [4].

Table 1: Conventional kernel functions.

Name	Kernel function expression
Linear kernel	$k(x_i, x_j) = x_i^T x_j$
Polynomial kernel	$k(x_i, x_j) = (t + x_i^T x_j)^d$
RBF kernel	$k(x_i, x_j) = exp(-\|\| x_i - x_j \|\|^2 = \sigma^2)$
MLP kernel	$k(x_i, x_j) = tanh(\beta_0 x_i^T x_j + \beta_1)$

D. SVM Algorithm

Assuming a linearly separable dataset, the task of learning coefficients w and b of support vector machine $f(\mathbf{x}) = (\mathbf{w}^T \mathbf{x}_i + b)$ reduces to solving the following constrained optimization problem:

$$\text{Minimize} \quad \tfrac{1}{2}\|\mathbf{w}\|^2$$

Subject to constraints

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i$$

This optimization problem can be solved by using the Lagrangian function defined as:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1]$$

, such that $\alpha_i \geq 0, \forall i$

where $\alpha_1, \alpha_2, \ldots \alpha_N$ are Lagrange multipliers and $\alpha = [\alpha_1, \alpha_2, \ldots \alpha_N]^T$.

The solution of the original constrained optimization problem is determined by the saddle point of $L(\mathrm{w}, b, \alpha)$ which has to be minimized with respect to w and b and maximized with respect to α.

Lagrange multipliers:

- If $y_i(\mathbf{w}^T \mathbf{x}_i + b) > 1$, the value of α_i that maximizes $L(\mathrm{w}, b, \alpha)$ is $\alpha_i = 0$.
- If $y_i(\mathbf{w}^T \mathbf{x}_i + b) < 1$, the value of α_i that maximizes $L(\mathrm{w}, b, \alpha)$ is $\alpha_i = +\infty$. However, since w and b are trying to minimize $L(\mathrm{w}, b, \alpha)$, they will be changed in such a way to make $y_i(\mathbf{w}^T \mathbf{x}_i + b)$ at least equal to +1.
- From this brief discussion, the so-called Kuhn_Tucker Conditions follow

$$\alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\} = 0, \quad \forall i$$

Data points x_i with $\alpha_i > 0$ are called the support vectors

Optimality conditions:

The necessary conditions to the saddle point of $L(\mathrm{w}, b, \alpha)$ are

$$\frac{\partial L}{\partial w_j} = 0, \quad \forall j$$

$$\frac{\partial L}{\partial \alpha_i} = 0, \quad \forall i$$

or stated a different way, $\nabla_{\mathbf{w}} L = 0, \quad \nabla_{\boldsymbol{\alpha}} L = 0$

Solving for the necessary conditions result in

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$$

$$\sum_{i=1}^{N} a_i y_i = 0$$

By replacing $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$ into the Lagrangian function and by using $\sum_{i=1}^{N} a_i y_i = 0$ as a new constraint the dual optimization problem can be constructed as

α that maximizes

$$\sum_i \alpha_i - \tfrac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \forall i$$

This is a convex quadratic programming problem, so there is a global minimum. There are a number of optimization routines capable of solving this optimization problem. The optimization can be solved in $O(N^3)$ time (cubic with the size of training data) and in linear time in the number of attributes.

E. Dataset

IRIS is R's own dataset. In this dataset there are 150 observations and 5 variables for each observation. The attributes are Sepal Length and Width, Petal Length and Width and Species.

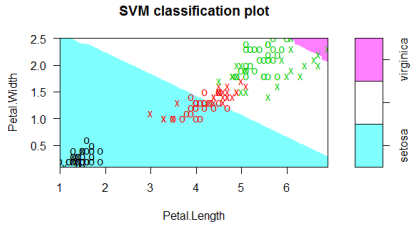
III. SVM PARAMETER TUNING USING GRID SEARCH METHOD

Grid search method is a complete and luxurious method which exhaustively generates candidate from a grid of parameter values with the specified range [5]. Using the default settings, the Grid Search Method only allows the join points to occur exactly at the observations. The maximum number of join points actually depends on the settings for number of observations which does not find the best fit. So a better fit can be achieved by using a finer grid by changing the setting for the number of data points to place between adjacent observed values in the grid search. So the Grid Search method creates a grid of all possible locations for join points specified by the settings, and calculates the Sum of Squared Error (SSE) at each one to find the best possible fit [6]. This method is computationally more efficient. The SVM algorithm is implemented on the IRIS dataset for multi-class classification in R which is a statistical tool. The package e1071 for SVM is installed in R using the keyword as given below:

```
> install.packages('e1071', dependencies=TRUE)
```

An SVM model is built using Species which is an attribute in the dataset to be used as instance classes. To view the built SVM model with a scatter plot of the input, the plot function can be used as follows:

```
> library(e1071)
> mymodel <- svm(Species~., data=iris,
kernel="sigmoid")
> summary(mymodel)

Call:
svm(formula = Species ~ ., data = iris, kernel =
"sigmoid")

Parameters:
 SVM-Type:  C-classification
 SVM-Kernel:  sigmoid
 cost:  1
 gamma:  0.25
 coef.0:  0

Number of Support Vectors:  54

 ( 6 26 22 )

Number of Classes:  3

Levels:
 setosa versicolor virginica
> plot(mymodel, data=iris,
Petal.Width~Petal.Length,slice=list(Sepal.Width=3,Sep
al.Length=4))
```

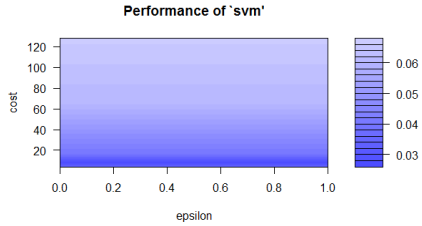


The predict function predicts values based on the model trained by SVM. It returns a vector of predicted labels for a classification problem. A cross-tabulation of the actual versus the predicted value results is called the confusion matrix.

```
> pre<-predict(mymodel, iris)
> tab<-table(Predicted = pre, Actual = iris$Species)
> tab
          Actual
Predicted    setosa versicolor virginica
  setosa         49          0         0
  versicolor      1         41         7
  virginica       0          9        43
> 1-sum(diag(tab))/sum(tab)
[1] 0.1133333
```

The tune function can be used to tune the hyperparameters of statistical methods using a grid search over the supplied parameter ranges [7]. This command will give the best parameters and best performance values.

```
> set.seed(123)
> tmodel<-tune(svm, Species~., data=iris, ranges =
list(epsilon = seq(0,1,0.1), cost = 2^(2:7)))
> plot(tmodel)
```




```
> summary(tmodel)
```

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
 epsilon cost
       0   8

- best performance: 0.02666667

- Detailed performance results:
   epsilon cost   error   dispersion
1   0.0   4 0.03333333 0.04714045
2   0.1   4 0.03333333 0.04714045
3   0.2   4 0.03333333 0.04714045
4   0.3   4 0.03333333 0.04714045
5   0.4   4 0.03333333 0.04714045
6   0.5   4 0.03333333 0.04714045
7   0.6   4 0.03333333 0.04714045
8   0.7   4 0.03333333 0.04714045
9   0.8   4 0.03333333 0.04714045
10   0.9   4 0.03333333 0.04714045
11   1.0   4 0.03333333 0.04714045
12   0.0   8 0.02666667 0.04661373
13   0.1   8 0.02666667 0.04661373
14   0.2   8 0.02666667 0.04661373
15   0.3   8 0.02666667 0.04661373
16   0.4   8 0.02666667 0.04661373
17   0.5   8 0.02666667 0.04661373
18   0.6   8 0.02666667 0.04661373
19   0.7   8 0.02666667 0.04661373
20   0.8   8 0.02666667 0.04661373
21   0.9   8 0.02666667 0.04661373
22   1.0   8 0.02666667 0.04661373
23   0.0   16 0.04000000 0.05621827
24   0.1   16 0.04000000 0.05621827
25   0.2   16 0.04000000 0.05621827
26   0.3   16 0.04000000 0.05621827
27   0.4   16 0.04000000 0.05621827
28   0.5   16 0.04000000 0.05621827
29   0.6   16 0.04000000 0.05621827
30   0.7   16 0.04000000 0.05621827
31   0.8   16 0.04000000 0.05621827
32   0.9   16 0.04000000 0.05621827
33   1.0   16 0.04000000 0.05621827
34   0.0   32 0.04666667 0.07062333
35   0.1   32 0.04666667 0.07062333
36   0.2   32 0.04666667 0.07062333
37   0.3   32 0.04666667 0.07062333
38   0.4   32 0.04666667 0.07062333
39   0.5   32 0.04666667 0.07062333
40   0.6   32 0.04666667 0.07062333
41   0.7   32 0.04666667 0.07062333
42   0.8   32 0.04666667 0.07062333
43   0.9   32 0.04666667 0.07062333
44   1.0   32 0.04666667 0.07062333
45   0.0   64 0.06000000 0.07981460
46   0.1   64 0.06000000 0.07981460
47   0.2   64 0.06000000 0.07981460
48   0.3   64 0.06000000 0.07981460
49   0.4   64 0.06000000 0.07981460
50   0.5   64 0.06000000 0.07981460
51   0.6   64 0.06000000 0.07981460
52   0.7   64 0.06000000 0.07981460
53   0.8   64 0.06000000 0.07981460
54   0.9   64 0.06000000 0.07981460
55   1.0   64 0.06000000 0.07981460
56   0.0 128 0.06666667 0.08314794
57   0.1 128 0.06666667 0.08314794
58   0.2 128 0.06666667 0.08314794
59   0.3 128 0.06666667 0.08314794
60   0.4 128 0.06666667 0.08314794
61   0.5 128 0.06666667 0.08314794
62   0.6 128 0.06666667 0.08314794
63   0.7 128 0.06666667 0.08314794
64   0.8 128 0.06666667 0.08314794
65   0.9 128 0.06666667 0.08314794
66   1.0 128 0.06666667 0.08314794

The cost parameter of the SVM model captures constraint violation. The default value of this parameter is 1. If cost is too high it will mean higher penalty for non-separable points and the model may share too many support vectors and it will lead to over fitting, whereas if the cost is too low it may end up with under fitting and the model may not be very accurate so that fixing high range of 2^2 to 2^7 of cost values for optimal solution.

IV. RESULTS AND DISCUSSION

Experiments were carried out on Iris dataset. The RBF kernel has better performance than the other kernels. After tuning the hyperparameters of the SVM model using the grid search method, the best parameters and best performance values are shown below. The sampling method used for performance evaluation is 10-fold cross validation. The details of the best parameter values are given below.

```
> mymodel<-tmodel$best.model
> summary(mymodel)

Call:
best.tune(method = svm, train.x = Species ~ ., data =
iris, ranges = list(epsilon = seq(0, 1, 0.1), cost =
2^(2:7)))

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  8
      gamma:  0.25


Number of Support Vectors:  35

 ( 6 15 14 )
```



```
Number of Classes:  3
Levels:
 setosa versicolor virginica
> plot(mymodel, data=iris, Petal.Width~Petal.Length,
slice=list(Sepal.Width=3,Sepal.Length=4))
```

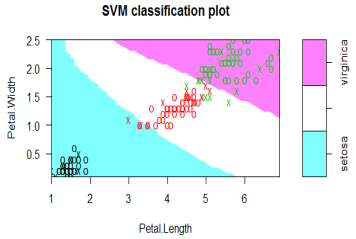


Fig. 2. Classification Plot after Parameter Tuning.

The confusion matrix before tuning the hyperparameters of SVM model is shown in Table2. The classification results were evaluated by means of accuracy. Accuracy of the SVM classifier before tuning the hyperparameters is 88.67% and its misclassification error is 11.33%.

Table 2: Before Tuning SVM Parameters.

	Actual		
Predicted	setosa	versicolor	virginica
setosa	49	0	0
versicolor	1	41	7
virginica	0	9	43

Table 3: After Tuning SVM Parameters.

	Actual		
Predicted	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	0
virginica	0	2	50

So the accuracy of the classifier after tuning SVM parameters has improved to 98.7% as compared to the accuracy of the classifier before tuning the parameters of the SVM model and the misclassification error rate is 1.33% shown in Table 3.

V. Conclusion

The traditional way of optimizing hyperparameters is a grid search method which is normally measured by cross validation on the dataset. A typical SVM classifier equipped with kernel has at least two parameters that require to be tuned for good performance. Since grid searching is a complete and possibly luxurious method, several alternatives have been proposed. In particular, Bio-Inspired algorithm with a randomized search that simply samples parameter settings a fixed number of times has been found to be more effective in high-dimensional spaces than grid search. This turns out some hyperparameters do not significantly affect the loss. Randomly dispersed data gives more textured data than grid search over parameters that ultimately do not affect the loss [8]. Further enhancements can be performed by using evolutionary algorithms for hyperparameter optimization.

REFERENCES

[1]. https://en.wikipedia.org/wiki/Unsupervised_learning.
[2]. Vladimar Vapnik, Statistical learning theory. Wiley, New Tork (1998).
[3]. https://machinelearningmastery.com/support-vector-machines-for-machine learning/.
[4]. Omid Naghash Almasi, Modjtaba Rouhani, "A new fuzzy membership assignment and model selection approach based on dynamic class centers for fuzzy SVM family using the firefly algorithm", *Turkish Journal of Electrical Engineering & Computer Sciences.*, Vol. **24**, Pp.1797 – 1814, 2016.
[5]. http://scikit-learn.org/stable/modules/grid_search.html.
[6]. https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/method/method-grid-search-or-hudsons.
[7]. https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Classification/SVM.
[8]. Bergstra, James; Bengio, Yoshua . "Random Search for Hyper-Parameter Optimization" *J. Machine Learning Research.*, Pp. 281–305, 2012.

International Journal of Theoretical & Applied Sciences, *Special Issue* 10(1a): 73-76(2018)
International Conference on e-SMAC-2018

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

A Comparative Study of Wireless Networks and Wireless Sensor Networks

P. Monika[1]and Dr. G. Kalpana[2]
[1]M.Phil. Research Scholar, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women Coimbatore (Tamilnadu), India,
[2]Associate Professor, Department of Computer Science,
Sri Ramakrishna College of Arts and Science for Women Coimbatore (Tamilnadu), India,

ABSTRACT: Wireless networking is used for providing the wireless connection between different network nodes. Wireless Sensor Networks are similar to wireless ad hoc networks where it can be used for the transmission of sensor data between the nodes. Sensing is the type of a technique helpful for gathering the information from the devices or processes. This paper mainly focuses on the characteristics of wireless and wireless sensor networks and their differences between. Wireless network and wireless sensor network faces some of the issues when the data transmission take place in the wireless medium.

Keywords: Wireless networks, Wireless sensor networks, Ad hoc networks, Multi-hop transmission, bandwidth, wireless medium.

I. INTRODUCTION

Wireless networking is used for providing the wireless connection between different network nodes. It can be used in many places like homes, offices, etc and it can be used for avoiding the process of investing more on cabling infrastructure. Some of the examples of wireless networks are cell phone network, wireless local area network, satellite communication, etc [1]. IEEE 802.11 introduced the Wi-Fi wireless technology for providing the internet connections for the users at any place. Wireless Local Area Network (WLAN) devices are used in many organizations with the stable network connection when the devices move from one place to another within the office.

Wireless Sensor Networks (WSN) are similar to wireless ad hoc networks where it can be used for the transmission of sensor data between the nodes. Wireless Sensor Networks are defined as the infrastructure-less wireless network which is used for monitoring the physical and environmental conditions like temperature, pressure, vibration, etc.WSN havemany functionalities like collecting, processing and transmitting those data to other devices. Zigbee is a type of WSN technology used for the transmission of data from one device to another through radio waves with low energy consumption [2].

II. WIRELESS NETWORKS

Wireless networking is used for providing the wireless connection between different network nodes [1]. In the wireless network technology, wireless links are required for transmitting the packets. The messages can be sent from the sender device via the wireless medium to the receiver within the transmission range of the sender. Wireless networks can be classified into two different types of network namely Infrastructure network and Ad Hoc networks.

Fig. 1. Wireless Networking Design.

In the infrastructure network, the access points (AP) are required for accessing the network. It can be denoted as the fixed technology in which the devices can connect to other device via the fixed device named as base station or access point [3]. Wireless network design with access points is represented in Fig 1. Ad Hoc Networks does not require the access points for the communication purpose. The nodes that are in the ad hoc network acts as the host or router. The devices can directly connect to other devices that are within the transmission range. When the sender tries to send message to the device that is not in coverage region then the message will be sent to the device within the sender transmission range and then sent to the exact receiver [4].

A. Characteristics of Wireless Networks

Some of the characteristics of wireless networks are discussed below:

Autonomous Behaviour: In the ad hoc networks, the devices can acts as a host or router, so that it has the capacity to switch the functionalities and work accordingly.

Multi-hop transmission: When the sender sends the message to the receiver out of the transmission range, then the message will be sent via the intermediate nodes [5].

Symmetric Environment: Since the devices in the wireless acts as a host or router, then all the devices have similar features and responsibilities, so that the devices are symmetric in nature.

Light weight features: The ad-hoc network devices have low CPU processing capability, less memory and power storage [4].

III. WIRELESS SENSOR NETWORKS

Wireless Sensor Networks (WSN) are similar to wireless ad hoc networks where it can be used for the transmission of sensor data between the nodes [2]. Sensing is the type of a technique helpful for gathering the information from the devices or processes. The main functionality of WSN is to monitor the environment and provide the report to the base station for further processing. WSN devices communicate with other devices via the gateway [6]. The gateway acts as a bridge between the WSN and other network. Wireless sensor network design is represented in Fig 2.

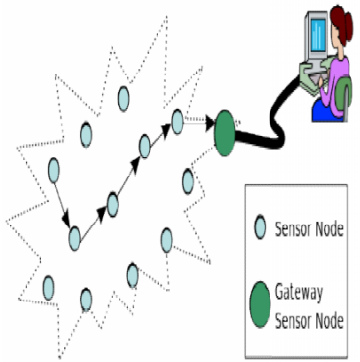


Fig. 2. Wireless Sensor Network.

The base station is a device that is used for controlling the activities of all other devices and take decisions according to the conditions [7]. Base stations play the vital role in wireless sensor network because the sensor nodes can cause many issues to the devices so that the base stations are used for collecting the data and take frequent backups of data. Those backup data can be used by the devices when the master node fails [8].Wireless sensor nodes are provided with many different sensors and those sensors can communicate between the devices using the radio signals. A wireless sensor node is constructed using the computing devices, radio transceivers and sensing devices. The sensor nodes are equipped with the built-in processor and it is used for processing the raw data before the transmission stage [9].

A. Characteristics of Wireless Sensor Networks

Characteristics of wireless sensor networks are explained below:

Scalability: Any number of nodes or devices can be attached to the infrastructure.

Responsiveness: The devices in the network are able to adapt to the changes that are made in future.

Reliability: The basic need of every network is to be reliable. The reliability is required in the data transmission of network structure.

Power efficiency: The devices in WSN have the ability to change its data path and handle all the nodes. It is designed in a way to work in power source other than direct power [10].

IV. WIRELESS NETWORKS Vs WIRELESS SENSOR NETWORKS

Features	**Wireless Networks**	**Wireless Sensor Networks**
Number of nodes	**Limited**	**Unlimited**
Sensing	**It does not posses.**	**It posses.**
Devices	**Devices like laptops, PDA's are used.**	**Very small devices are used.**
Power	**Limited capacity**	**Larger capacity**
Cost	**Expensive**	**Inexpensive.**
Transmission range	**Ranges from 10 – 500 meters.**	**Ranges from 3 – 30 meters.**
Node density	**Lower**	**Higher**
Communication	**Data moves from one device to many devices.**	**Data moves from many devices to the gateway.**

V. CHALLENGES IN WIRELESS NETWORKS

Some of the key challenges of wireless networks are:

A. Signal fading

In the wireless network infrastructure the signals are transmitted over the wireless transmission medium, so that the signals can travel in different path due to reflection or scattering caused by other objects. When the signals are transmitted in different directions it takes different time duration to reach the destination. Once the receiver get the signals it may has more

noise or attenuation than the original transmitted signal. Thus signal fading issue in wireless network cause more packet loss.

*B. Power and Energy***:**
The devices in the wireless networks can be smaller in size and they are dedicated to perform some set of functions, but the power resources cannot be shared to the devices as expected. When the device moves from one place to other, it is hard to provide continuous power supply to all the movable devices.

C. Data Rate
Data rate is the major issue in wireless networks because the multimedia applications like audio and video should be supported by the devices with high speed data rate. the comparison ratios of the multimedia applications can be between 75 to 100. So that the audio and video with good quality should be provided at this comparison rates [3].

D. Security
Security is also a major issue in wireless networks because the data are transmitted via the radio frequencies. When the data are sent via wireless medium, the intruders have the access to the data that is sent. They can able to steal the data or deny the service to the users or receivers [1].

VI. CHALLENGES IN WIRELESS SENSOR NETWORKS

Some of the key challenges of wireless sensor networks are:

A. Limited Bandwidth
In WSN less power will be consumed for data processing when compared to data transmission. During data transmission, only limited bandwidth will be used so that it affects the messages that are sent between the sensors.

B. Energy
Power consumption in wireless senor devices are allotted to three domains namely sensing, communication and data processing. All these factors depend on the battery life of the devices. Usually, the batter life can be recharged or replaced when it get decreased. But for some applications the sensors battery life should be available until its work has been completed [8].

C. Security
There are more security issues in the wireless sensor networks. Data integrity is the major issue in WSN. Since the data are sent via transmission medium, the receivers should ensure that the data they have received is original. The data that are sent should be authenticated and it should not be modified by the unknown users [11].

VII. APPLICATIONS

The applications that are used for wireless networks and wireless sensor networks are discussed below:

A. Wireless Networks
Military battlefield: Military equipments are provided with some of the computer equipments. It helps to maintain information among the vehicles, soldiers and military head quarters.
Commercial sector: Wireless network devices can be used in emergency or rescue operations for natural calamities reliefefforts, e.g. in fire, flood, or earthquake. Information is delivered from one rescue team member to another.
*Local level:*Wirelessdevices can communicate between instant and temporary multimedia networkusing notebook computers or palmtop computers to spread and share information among participants ata conference. Another appropriate local level application might be in home networks where devicescan communicate directly to exchange information [4].

B. Wireless Sensor Networks
Area monitoring applications: In area monitoring, the WSN can be deployed in a region to monitor any activities. When the sensors detect the event being monitored, the event is reported to the base station, which then takes appropriate action.
Environmental applications: A few environmental applications of sensor networks include forest fire detection, green house monitoring, etc. They can be used for tracking the movement of insects, birds and small animals.
*Health applications:*Some of the health applications for sensor networks are providing interfaces for tracking and monitoring doctors and patients inside a hospital and telemonitoring of human physiological data, etc [11].

VIII. CONCLUSION

Wireless technologies are used in many organizations in order to reduce time and the issues that are in the cables of wired infrastructure. This paper mainly focuseson characteristics, issues and applications of both wireless networks and wireless sensor networks.

REFERENCES

[1]. Surabhi Surendra Tambe,"Wireless Technology in Networks", *International Journal of Scientific and Research Publications*, ISSN 2250-3153, Vol. **5** Issue 9, July 2015.
[2]. Aamir Shaikh and Siraj Pathan, "Research on Wireless Sensor Network Technology", *International Journal of Information and Education Technology*, Vol. **2** Issue 5, Oct 2012, pp 476-479.
[3]. Aniruddha Singh, Abhishek Vaish and Pankaj Kumar Keserwani, "Research Issues and Challenges of Wireless Networks", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Vol 4 Issue 2, Feb 2014, pp 572-575.

[4]. Jagtar Singh and Natasha Dhiman, "A Review Paper on Introduction to Mobile AdHoc Networks", *International Journal of Latest Trends in Engineering and Technology*, ISSN: 2278-621X, Vol. **2** Issue 4, July 2013, pp 143-149.
[5]. Sushmita Kopekar and Amresh Kumar, "A Study of Ad-Hoc Wireless Networks: Various Issues in Architectures and Protocols", *International Journal of Computer Applications*, ISSN 0975 – 8887, Vol. **122** Issue 6, July 2015, pp 36-40.
[6]. Murat Dener, "A new gateway node for wireless sensor network application", *Scientific Research and Essays*, ISSN 1992-2248, Vol. **11** Issue 20, Oct 2016, pp 213-220.
[7]. Shantala Devi Patil and Vijayakumar B P, "Overview of Issues and Challenges in Wireless Sensor Networks", International Journal of *Application or Innovation in Engineering & Management*, ISSN 2319 – 4847, Vol. **5** Issue 5, May 2016.
[8]. S. Karthik and Dr. A. Ashok Kumar, "Challenges of Wireless Sensor Networks andIssues associated with Time Synchronization", *International Journal of Advanced Networking and Applications*, Mar 2015, pp 19-23.
[9]. Edwin Prem Kumar Gilbert, Baskaran Kaliaperumal, and Elijah Blessing Rajsingh, "Research Issues in Wireless Sensor Network Applications: A Survey", *International Journal of Information and Electronics Engineering*, Vol. **2** Issue 5, Sep 2012, pp 702-706.
[10]. Muhammad R Ahmed, Xu Huang, Dharmandra Sharma and Hongyan Cui, "Wireless Sensor Network: Characteristics and Architectures", *International Journal of Information and Communication Engineering*, Vol. **6** Issue 12, 2012, pp 1398-1401.
[11]. Himani Chawla, "Some issues and challenges of Wireless Sensor Networks", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Vol. **4** Issue 7, July 2014, pp 236-239.

International Journal of Theoretical & Applied Sciences,

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

INSTRUCTIONS TO AUTHORS

The journal invites full and mini-review research articles, full length research articles and brief communication giving exciting current information of all Applied Sciences, Pharmaceutical Sciences.

Title and authors names and address must be given in the front page. This must be followed by the abstract on separate page. Thereafter, material and method, results and discussion, acknowledgements, references. Tables, photographs and all figures including drawings, graphics, and diagrams must be attached after the references. Five type of manuscript may be submitted like, **Original Papers, Rapid communication, Technical note, Review, Advancements in instrumentation.** All manuscripts should be typed double-spaced and in 10 pt Time New Roman including, references, tables, figure). Manuscripts preferred for publication contain original work, focused on the core aims and scope of journal, clearly and correctly written and should be written clearly and grammatically. The manuscript must be accompanied by a cover letter. It should have 50-100 words that contains Significance of work, Novelty of the work, Contribution to field/community.

Paper elements

Title page with: Names of authors with address and Personal e-mail addresses and mobile numbers

Abstract, Keywords, Introduction, Material and Methods, Results and Discussion, **Acknowledgments, Reference lists**,

Tables, Figure captions Figures should be adjust in text where they needed.

Title of article – Title should be concise and informative describing the contents of pages.

Author name and affiliation – Author(s) name should be consists of first and middle name initial followed by full last name. In case with more than one author clearly indicate who is willing to handle correspondence concerning the paper i.e. correspondence author and use asterisk to indicate the corresponding author. Use the symbol *, **,***, **** etc as superscripts to relate the authors to corresponding addresses. Provide the e-mail address, full postal address, mobile numbers of each author.

Abstract – The abstract should present a brief summary of the paper including questions being addressed and the key findings of the study. It should not serve as an introduction nor contain references. It consist of one paragraph not exceed 200 words.

Keywords – It provide immediately after the abstract and between 3-10 keywords. Keywords assist readers for indexing purposes.

Introduction – It include the scientific importance, historical background relevance to other area and objectives of the paper.

Material and Methods – It should be written in sufficient detail to enable others to repeat the author(s) work.

Results and Discussion - It may be combined or kept separate and may be further divided into subsections. This section should not contain technical details.

Acknowledgments – Acknowledgments of people, grants, funds etc. should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

Tables and Figure – Tables and figures should not be embedded in the text, but should be included as a separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. Avoid vertical rules. They are numbered consecutively in accordance with their appearance in the text. Number the tables as Table 1, Table 2 etc, to in the text. Figures should be completely labeled, taking into account necessary size reduction.

References - Within the text, references should appear as consecutive numbers in brackets (e.g., [1], [1,2,3]). The list of references should be given in the order of the first appearance of references in the text. The list of references should be formatted as follows:

1. For Articles in Journals: Indicate the initials and surnames of the authors, the title of the journal in italic, the volume number, the number of the first page, the year of the reference (in parentheses), for example,

M. Smith, G. Gaur, and I. Mehta, *Laser Phys.*, **1**, 123 (1991).

2. For Books: The initials and surnames of the authors, the full title of the book in italic, and (in parentheses) the publisher, the city, and the year, for example,

S.S. Mishra, S. Morgan, and P. Charlton, *Quantum Mechanics* (Allen, New York, 1990).

3. Conference Proceedings: Please add all available data such as title, date, and place of the conference as well as publisher, place, and year of publication or, alternatively, for example,

J. Ansell, I. Harrison, and C. T. Foxon: Proceedings of the 4th International Conference on Chemistry, Colorado, USA, 2001, Part A (Wiley-VCH, Berlin, 2002), pp. 279-282.

4. References to Online Material: Should include a brief description and or title:

http://www.kzoo.edu/ajp/docs/information.html.

5. Reference to a Thesis: Author initial, surname, DSc/PhD/MSc/BSc thesis, university, town, country and year of publication in bracket.

A.J. Agutter, Ph.D thesis, Edinburgh University (Edinburgh, UK, 1995).

Equations - Each equation should appear on a separate line with proper punctuation placed before and after it. All equations must be numbered sequentially. The number of the equation, in parentheses, should be placed near the right-hand margin. Avoid bars either above or below letters. Avoid subscripts on subscripts, etc. Adequate space must be allowed for marking of inferior and superior letters or numbers. Crowded equations lead to errors in composition. Use the following format to refer to equations in the text: Equation (5) follows from substituting Eqs. (2) and (3) into Eq. (4).

Chemical Reaction Data - *(Relevant for only Chemical and Biochemical Fields*

For heterogeneous catalysis, presentation should include reaction rates normalized by catalyst surface area, surface area of the active phase, or number of active surface atoms or catalytic sites, as appropriate. Typical rate units are mol s^{-1} m^{-2} or, in the case of surface atom normalization to produce turnover frequencies, s^{-1}. For homogeneous catalysis, rates should typically be reported as turnover frequencies. Comparisons of selectivities should be made at similar conversions. Catalytic measurements need to be carried out under kinetically limited conditions. Confirming tests need to be carried out and reported, especially for all reactions occurring in the liquid phase.

Symbols and Units- Greek symbols and special characters often undergo formatting changes and get corrupted or lost during preparation of a manuscript for publication.

To ensure that all special characters used are embedded in the text, these special characters should be inserted as a symbol but should not be a result of any format styling (*Symbol* font face) otherwise they will be lost during conversion to PDF/XML^2. Authors are encouraged to use SI units, but use of SI units is not mandatory if other units are more appropriate.

Nomenclature - Nomenclature should confirm to current American usage. Insofar as possible, authors should use systematic names similar to those used by Chemical Abstracts Service or IUPAC.

External review - The external review process is initiated when the editor sends the manuscript out for review. When the reports are returned, the Editor makes a decision based on the recommendations received and the number of previous revisions and informs the submitter in a decision letter.

Decisions on initial submissions

a. If the reviews are negative or insufficiently strong to support continued editorial consideration, the manuscript will be rejected.
b. In other cases, including cases where the reviews are mixed, the manuscript will be returned for revision with suggestions and directions for resubmission.
c. With the resubmission, authors must include a cover letter that summarizes the revisions and provides responses to the issues and questions raised by the editor and/or the reviewers. Upon resubmission the manuscript will usually be sent back to the previous reviewers and occasionally to new reviewers for re-review.

Revised manuscript - A revised manuscript should be returned within 6 days for minor changes and 10 days for major revisions. If the manuscript is not returned within this time frame, it will be considered withdrawn by the author and any revised version submitted subsequently will be considered a new contribution.

After acceptance

Copyright transfer – Authors will be asked to transfer copyright of the article to the publisher. This will ensure the widest protection and dissemination of information under copyright laws.

Offprint – Additional offprint can be ordered by the corresponding authors.

Proof reading – The corresponding author will receive a proof and should be return to publisher with in three days of receipt. Correction should be restricted to typesetting error; any other correction may be checked and corrected since the inclusion of late correction cannot be accepted.