ISSN (Print):2321–404X, (Online):2321–4384

INTERNATIONAL JOURNAL OF SOFT COMPUTING & ARTFICIAL INTELLIGENCE

VOLUME NO. 11
ISSUE NO. 3
SEPTEMBER - DECEMBER 2023

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

ISSN (Print):2321–404X, (Online):2321–4384

International Journal of Soft Computing & Artificial Intelligence

Aim & Scope

IJSCAI provides a new forum for dissemination of knowledge on both theoretical and applied research on Artificial Intelligence and Soft Computing with an ultimate aim to bridge the gap between these two non-coherent disciplines of knowledge. The Journal offers survey and review articles from experts in the field, promoting insight and understanding of the state of art, and latest trends in the field. The content includes original research and innovation ideas, applications from all over the world.

All published papers are also available freely with online full-text content. This forum accelerates interaction between the above two bodies of knowledge, and fosters unified development in next generation computational models for machine intelligence. Special Issues devoted to topics of current interest will be published occasionally.

ISSN (Print):2321–404X, (Online):2321–4384

ISSN (Print):2321–404X, (Online):2321–4384

International Journal of Soft Computing & Artificial Intelligence

(Volume No. 11, Issue No. 3, September - December 2023)

Contents

ISSN (Print):2321–404X, (Online):2321–4384

“Data Mining Techniques Differentiation Between NPPLS and PLS”

Manoj Kumar
Department of Computer Science Govt. College, Narnaul
09416259830

ABSTRACT

This paper compares two different predictive data-mining techniques (one linear technique, Partial Least Squares (PLS) and one nonlinear technique(NLPLS)) on two different and unique data sets: a collinear data set (called "the COL" data set in this paper) and a simulated data set (called "the Simulated" data in this paper). These data are unique, having a combination of the following characteristics: few predictor variables, many predictor variables, highly collinear variables, very redundant variables and presence of outliers. The natures of these data sets are explored and their unique qualities defined. This is called data pre-processing and preparation. To a large extent, this data processing helps the miner/analyst to make a choice of the predictive technique to apply. The big problem is how to reduce these variables to a minimal number that can completely predict the response variable. the Partial Least Squares (PLS, a supervised technique), and the Nonlinear Partial Least Squares (NLPLS), which uses some neural network functions to map nonlinearity into models, were applied to each of the data sets . Each technique has different methods of usage; these different methods were used on each data set first and the best method in each technique was noted and used for global comparison with other techniques for the same data set. The purpose of this is to identify the technique that performs best for a given type of data set and to use it directly instead of relying on the usual trial-and-error approach. When this process is effectively used, it will reduce the lead time in building models for predictions or forecasting for business planning. The work in this Research paper will also be helpful in identifying the very important predictive data-mining performance measurements or model evaluation criteria.

Keywords: PLS, NLPLS, COL, PDM

INTRODUCTION

Nearly all areas of life activities demonstrate a similar pattern. Whether the activity is finance, banking, marketing, retail sales, production, population study, employment, human migration, health sector, monitoring of human or machines, science or education, all have ways to record known information but are handicapped by not having the right tools to use this known information to tackle the uncertainties of the future. Breakthroughs in data-collection technology, such as bar-code scanners in commercial domains and sensors in scientific and industrial sectors, have led to the generation of huge amounts of data [1]. This tremendous growth in data and databases has spawned a pressing need for new techniques and tools that can intelligently and automatically transform data into useful information and knowledge. For example, NASA's Earth Observing System, which is expected to return data at the rate of several gigabytes per hour by the end of the century, has now created new needs to put this volume of information to use in order to help people make better choices in that area [2]. These needs include the automatic summarization of data, the extraction of the “essence” of information stored, and the discovery of patterns in the raw data. These can be achieved through data analyses, which involve simple queries, simple string matching, or mechanisms for displaying data [3]. Such data-analysis techniques involve data extraction, transformation, organization, grouping, and analysis to see patterns in order to make predictions. To industrial engineers, whose work it is to devise the best means of optimizing processes in order to create more value from the system, data-mining becomes a powerful tool forevaluating and making the best decisions based on records so as to create additional value and to prevent loss. The potential of data-mining for industrial managers has yet to be fully exploited

.

1. PREDICTIVE DATA MINING:

Hence, with the advent of improved and modified prediction techniques, there is a need for an analyst to know which tool performs best for a particular type of data set. In this paper, two prediction tools Partial Least Squares [PLS]; and Nonlinear Partial Least Squares [NLPLS]), are used on two uniquely different data sets to compare the predictive abilities of each of the techniques on these different data samples. The advantages and disadvantages of these techniques are discussed also. Hence, this study will be helpful to learners and experts alike as they choose the best approach to solving basic data- mining problems. This will help in reducing the lead time for getting the best prediction possible.

Data mining is the exploration of historical data (usually large in size) in search of a consistent pattern and/or a systematic relationship between variables; it is then used to validate the findings by applying the detected patterns to new subsets of data [7, 8]. The roots of data mining originate in three areas: classical statistics, artificial intelligence (AI) and machine learning [9]. Pregibon [10] described data mining as a blend of statistics, artificial intelligence, and database research, and noted that it was not a field of interest to many until recently. According to Fayyad [11] data mining can be divided into two tasks: predictive tasks and descriptive tasks. The ultimate aim of data mining is prediction; therefore, predictive data mining is the most common type of data mining and is the one that has the most application to businesses or life concerns. DM starts with the collection and storage of data in the data warehouse. Data collection and warehousing is a whole topic of its own, consisting of identifying relevant features in a business and setting a storage file to document them. It also involves cleaning and securing the data to avoid its corruption. According to Kimball, a data ware house is a copy of transactional or non-transactional data specifically structured for querying, analyzing, and reporting [12]. Data exploration, which follows, may include the preliminary analysis done to data to get it prepared for mining. The next step involves feature selection and or reduction. Mining or model building for prediction is the third main stage, and finally come the data post-processing, interpretation, and/or deployment. Applications suitable for data mining are vast and are still being explored in many areas of business and life concerns. This is because, according to Betts [13], data mining yields unexpected nuggets of information that can open a company's eyes to new markets, new ways of reaching customers and new ways of doing business. For example, D. Bolka, Director of the Homeland Security Advanced Research Project Agency HSARPA (2004), as recorded by IEEE Security and Privacy [14], said that the concept of data mining is one of those things that apply across the spectrum, from business looking at financial data to scientists looking for scientific data. The Homeland Security Department will mine data from biological sensors, and once there is a dense enough sensor network, there will be enormous data flooding in and the data-mining techniques used in industries, particularly the financial industry, will be applied to those data sets. In the on-going war on terrorism in the world especially in the United States of America (after Sept. 11th of 2001), the National Security Agency uses data mining in the controversial telephone tapping program to find trends in the calls made by terrorists with an aim to aborting plans for terrorist activities. Table 2.1 is an overview of DM's applications. In the literature, many frameworks have been proposed for data-mining model building, and these are based on some basic industrial engineering frameworks or business improvement concepts. Complex data-mining projects require the coordinated efforts of various experts, stakeholders, or departments throughout an entire organization in a business environment; therefore, this makes needful some of the frameworksproposed to serve as blueprints for how to organize the process of data collection, analysis, results dissemination and implementing and monitoring for improvements.

1.1 Partial Least

The score vectors are the values of the data on the loading vectors p and q. Furthermore, a principle component-like analysis is done on the new scores to create loading vectors (p and q). Figure 2.4, an inferential design of PLS by Hines [15], is a representation of this. In contrast to principal component analysis (PCA), PLS focuses on explaining the correlation matrix between the inputs and outputs but PCA dwells on explaining the variances of the two variables. PCA is an unsupervised technique and PLS is supervised. This is because the PLS is concerned with the correlation between the input (x) and the output (y) while PCA is only concerned with the correlation between the input variables x. As can be seen in Figure 2.4, b would represent the linear mapping section between the t and u scores. The good point of PLS is that it brings out the maximum amount of covariance explained with the minimum number of components. The number of latent factors to model the regression model is chosen using the reduced eigenvectors.

The eigenvectors are equivalent to the singular values or the explained variation in the PC selection and are normally called the Malinowski's reduced eigenvalue [6]. When the reduced eigenvalues are basically equal, they only account for noise.

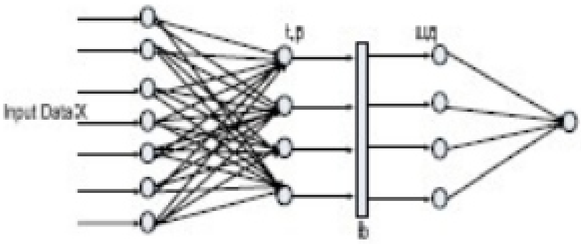


Figure1.1 diagram of the PLS Inferential Design.

1.2 Non Linear Partial Least Squares (NLPLS)

This is shown diagrammatically in Figure 2.5, an inferential design of NLPLS by Hines [15]. It is just the same as the process explained above, with the major difference being that in the linear PLS method, the inner relationships are modeled using simple linear regression. The difference between PLS and the NLPLS models is that in NLPLS, the inner relationships are modeled using neural networks [20,19]. For each set of score vectors retained in the model, a Single Input Single Output (SISO) neural network is required [15]. These SISO networks usually contain only a few neurons arranged in a two-layered architecture. The number of SISO neural networks required for a given inferential NLPLS unit is equal to the number of components retained in the model and is significantly less than the number of parameters included in the model [16].

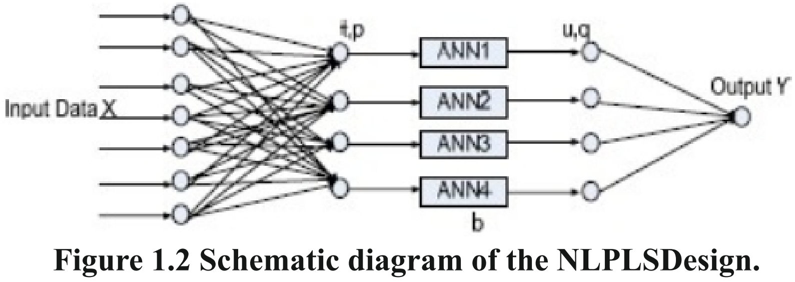


Figure 1.2 Schematic diagram of the NLPLSDesign.

PROCEDURE

The relationship check is made by plotting the inputs over the output of the raw data sets. The data is preprocessed by scaling or standardizing them (data preparation) to reduce the level of dispersion between the variables in the data set. The correlation coefficients of each of the various data sets are computed to verify more on the relationship between the input variables and the output variables. This is followed by finding the singular value decomposition of the data sets transforming them into principal components. This also will be helpful in checking the relationship between the variables in each data set. At this stage, the data sets are divided into two equal parts, setting the odd number data points as the "training set" and the even number data points as the "test validation data set." Now the train data for each data set is used for the model building. For each train data set, a predictive data mining technique is used to build a model, and the various methods of that technique are employed. For example, Multiple Linear Regression has three methods associated with it in this Research paper: the full model regression model, the stepwise regression method, and the model built selecting the best correlated variables to the output variables. This model is validated by using the test validation data set. Nine model adequacy criteria are used at this stage to measure the goodness of fit and adequacy of the prediction. The results are presented in tables. The results of the train sets are not presented in this study because they are not relevant. This is because only the performance of the model on the test data set or entirely different (confirmatory) data set is relevant. The model is expected to perform well when different data sets are applied to it. In this Research paper work, the unavailability of different but similar real-life data sets has limited this study to using only the test data set for the model validation. This is not a serious problem since this work is limited to model comparison and is not primarily concerned with the results after deployment of the model.

Finally, all the methods of all the techniques are compared (based on their results on each data set) using four very strong model adequacy criteria. The best result gives the best prediction technique or algorithm for that particular type of data set.

2. DATA INTRODUCTION:-

2.1 The COL Data Set Description and Preprocessing:-

The variables or attributes are simply designated as Variables 1 to 8, with variable 8 being the output variable and the rest being input variables. It was necessary to scale these to reduce the dispersion and bring all the variables to the same unit of measure. After scaling, the entire matrix now has a mean of 0 and standard deviation. The correlation coefficient matrix will reveal this relation better. It can be observed that all the variables are strongly correlated with each other. Indeed, they are almost perfectly correlated with each other and with the response variable. Therefore there is no nonlinear relationship between those PCs' scores plotted. This will further be revealed from the plot of inner scores matrix and outer score matrix of the Partial Least Square model and that of the Nonlinear Partial Least Squares.

2.2 The Simulated Data Set Description and Preprocessing:-

The last set of data used in this analysis is the simulated data. This data set has 44 variables and 5,000 data points. Variable 38 (Column 38) is the response variable and the rest are independent variables. The variables ranged from -5 to 20, with some spikes showing outliers or noise. The majority of the variables in the data have values above 5, so there is still need for standardization of the data set to reduce the degree of dispersion between the data points in the matrix. The data set was scaled once again to reduce this dispersion and give every point an equal opportunity of showing up in the matrix. Some of the variables showed good correlation with the output variable, but a great number of them didn't. After the scaled data were plotted, the cluster was now about zero (ranging from -2 to 2), with spikes showing outliers. The correlation coefficient matrix revealed a very weak correlation between the input and the output variables. One cannot see any presence of a nonlinear relationship, although the large data points may have hidden any trace of them.

2. THE STATISTICS OR CRITERIA USED IN THE COMPARISON:-

In this section, the entire nine criteria used to compare the various methods within each technique are briefly explained.

1. R-square (R2 or R-Sq) measures the percentage variability in the given data matrix accounted for by the built model (values from 0 to 1).
2. R-square Adjusted (R2 adj) gives a better estimation of the R2 because it is not particularly affected by outliers. While R-sq increases when a feature (input variable) is added, R2 adj only increases if the added feature has additional information added to the model. R2 adj values ranged from 0 to 1.
3. Mean Square Error (MSE). MSE measures the difference between the predicted test output and the actual test outputs. The smaller the MSE, the better. Large MSE values mean poor prediction.
4. Root Mean Square Error (RMSE); this is just the MSE in the units of the original predicted data. It is calculated by finding the square root of MSE.
5. Mean Absolute Error (MAE); this quantity takes care of overestimation due to outliers.
6. Modified Coefficient of Efficiency (E-mod); the modified coefficient of efficiency gives information equivalent to the MAE (values from -1 to 1).
7. The Weight of the regression models (norm); this value calculates the weights of the regression coefficients.
8. Condition number of the predictor matrix (CN); this quantity, designated as CN here, gives a measure of the stability of the model built. High condition numbers (> 100) show that the problem is ill- conditioned and hence cannot give consistent or stable results.
9. Number of features or variables used (N). The objective of every builder is to make use of the smallest amount of resources to achieve the desired result, as per Occam's Razor. Since data collection and analysis are expensive, fewer features (variables) take less energy and resources to deal with.

The data sets are divided into two: the training set and the test data sets (odd numbered data points are the training set and even-numbered data points are the test set). The train set predictor (input) variables are used to build the model. The train set predictor is regressed against the train response variable, and the resulting regression constants are called the regression coefficients. These coefficients are post-multiplied with the train set input variables to get the predicted train set response variable. This is called the training. When the prediction is compared with (subtracted from) the original train data, the difference between the prediction and the original output is revealed. When the same coefficients are post-multiplied with the test set input (predictor) variables, the result is also compared with the test set original response output. The test set is used to confirm the soundness of the model. The ability to accurately predict the test output tells how good the model is and is a measure of model performance. The results from the predicted training sets' output are important because a model is expected to perform well in the training set used to build the model. In this work however, the results were not included because of size and most importantly, in real-life analysis, the soundness of the model is only measured by its ability to predict new data sets and not the train data sets from where the model was trained. If a model performed very well in the training set and could not perform satisfactorily in the test validation data set or new data sets, then its predictive ability is suspect and cannot be used for prediction. Hence, the results of the predictions of the training sets' output were not presented in this analysis because the performances of the models in the test validation data set are of more significance to this study. Only the predictions of the response variables of the test validation data sets were used for the model comparison.

3. COL DATA SET ANALYSIS:-

The description of the COL data shows that the variables were almost perfectly correlated with each other; therefore this data set was divided in a unique way to avoid the replication of the train data set on the test data set. Dividing this data as before would mean having the train data set and test data set is almost the same, so in this case, the data were divided into blocks of 200s. The odd blocks were the training set, and the even blocks were the test set (this is different from the earlier division into odd numbers as training set and even numbers as test set). In this division, care was taken to make sure the training set covered the entire data matrix.

4.1 Partial Least Squares (PLS) on COL data:-

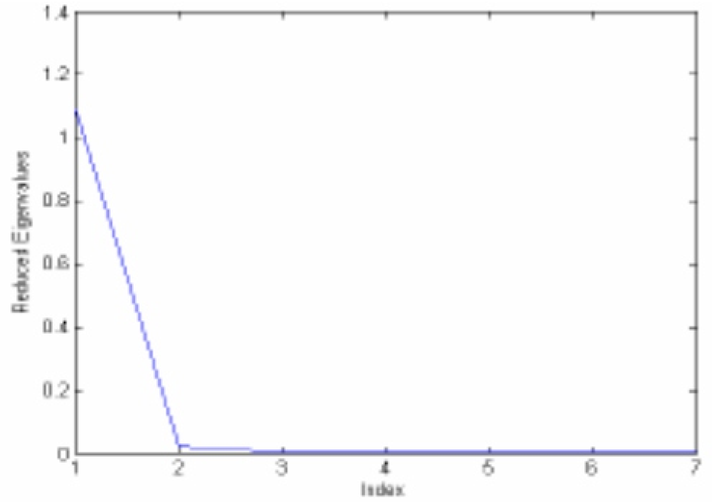
In the COL data analysis, three models of the PLS were built. Using Malinowski's eigenvalues, (Table 4.1), the plot of the reduced eigenvalues against the index shows that two factors looked significant (Figure 4.1). Using the iterative method, the minimum MSE gave four optimal numbers of factors (Figure 4.2). Finally a model was made using all the factors. The result of these various models is shown in the Summary Table (Table 4.2).

As can be seen in Table 4.2, the best model is the optimal eigenvalue model (four factors). From the reduced eigenvalues Table 4.1, the fifth factor to the seventh factor seemed to have reduced eigenvalues that were equivalent and could be classified as noise. The solution of the optimal factors (4 factors) and that of the model built with all the factors looked almost the same in terms of the R.Sq. the R2 adj, the RMSE, the MAE and the modified coefficient of efficiency. Their condition numbers CN were above 2,000. The model built with only two factors had a good condition number (49), and the R.Sq and R2 adj were not bad, but the MSE was relatively high (57.8). Figure 4.3 shows the predictions of the test output data with two, four and all of the seven factors. Figure 4.4 shows the output scores plotted over the input scores (predicted and test response). This plot shows that the model is a linear one. The data itself looked linear and the model represents that. The generalization of the linear pattern was good. Perhaps NLPLS after training will copy the nonlinearity or over-fit the data.

Figure 4.1 Plot of the reduced eigenvalues vs. the index.

Table 4.1 Malinowski's reduced eigenvalues for the COL data.

	Reduced Eigenvalues
1	1.092
2	0.0206
3	0.0007
4	0.0003
5	0.0001
6	0
7	0



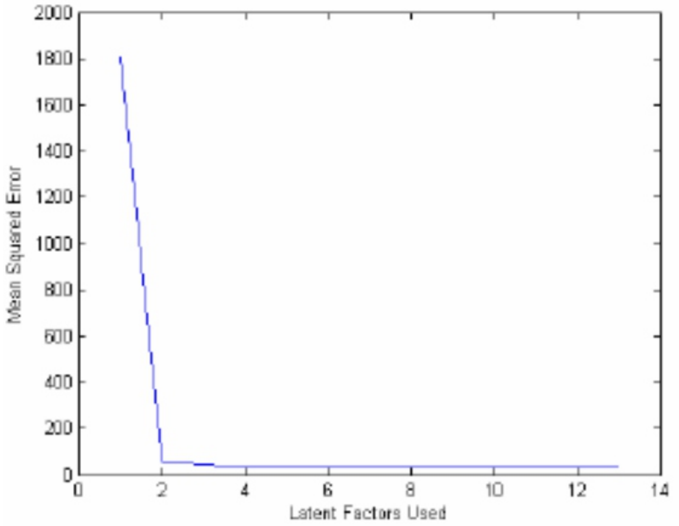
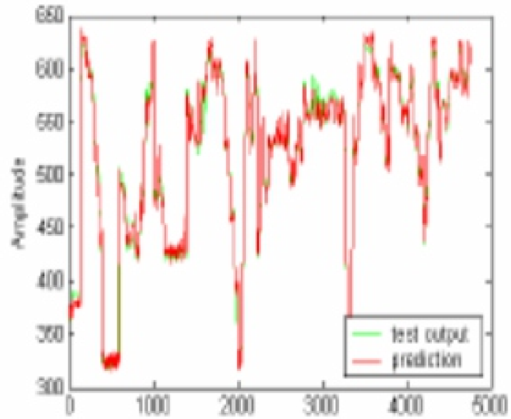


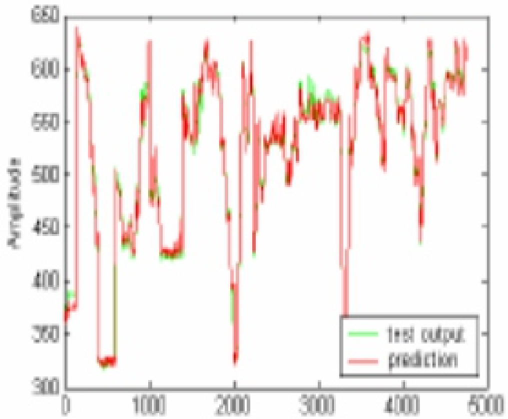
Figure 4.2 Plot of the Mean Square Error vs. the latent factors.

Table 4.2 Summary of the PLS results on the COL data set.

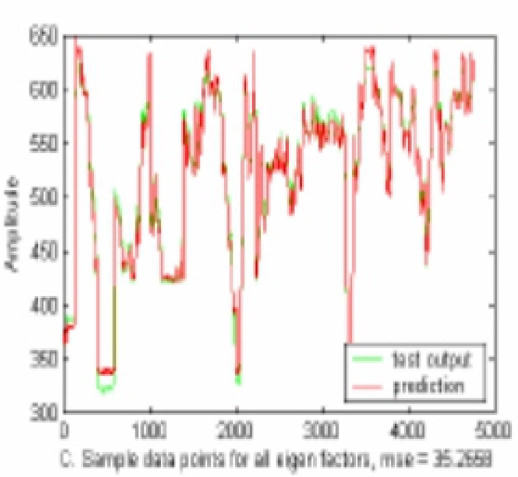
PLS	R-Sq	R-Sq-Adj	MSE	RMSE	MAE	E-mod.	CN	Normwt	N
Red. Eig. Val	0.9908	0.9908	57.8274	7.6044	5.8683	0.9094	49.13	1.0215	2
Optimal Val	0.9946	0.9946	33.9342	5.8253	4.651	0.9278	2311.4	1.029	4
All factors	0.9944	0.9944	35.2658	5.9385	4.7274	0.9266	8940.5	1.3288	7



A. Sample data points for for 2 eigen factors, mse = 57.8274



B. Sample data points for 4 eigen factors, mse = 33.9342



4.2 Non-Linear Partial Least Squares (NLPLS) on the COL Data

Two models were built with NLPLS. The first model was built with the optimal number of factors, and the second was built on [after?] retraining the data. The neural network training function was used to train the train data set until the performance goal was met. Using the iterative method, the minimum mean absolute error was computed and plotted against the latent factor (Figures 4.5 to 4.6). The results showed some inconsistencies. At the first training, the optimal latent factors value was 4 (Figure 4.5), at the second training it changed to 2 (Figure 4.6), and in the third training the optimal latent factors value was 5. These gave three different MAE (see table 4.3). The results of the three NLPLS models are given in Table 4.3. The model with only two factors outperformed the one with four factors. The optimal latent factors of 5(C) gave MSE 22.7752 and MAE of 3.4819. The solution was not stable with NLPLS and therefore was unreliable. It was observed that when the data were retrained, new optimal results emerged. This was repeated many times over, and different optimal results were obtained each time. When two similar optimum factors resulted, the statistics for model evaluation also differed. Figure 4.7 shows the plots of the output scores over the input scores for the predicted and the test response. It is very obvious that there is nonlinearity in the model. NLPLS also mapped the nonlinearity contained in the data.

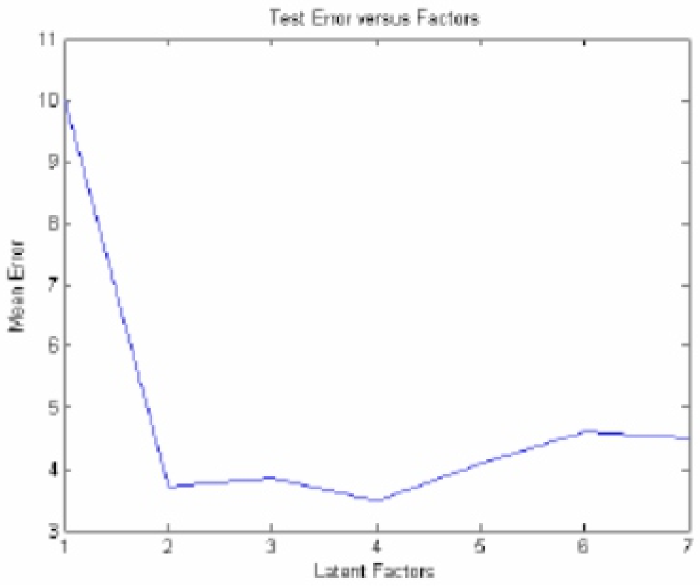


Figure 4.3Plot of the MAE against the latent factors after first neural network training.

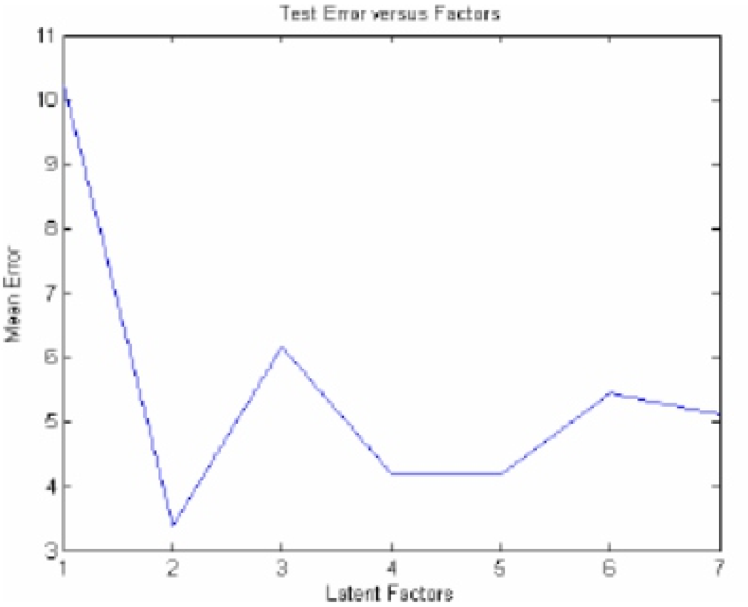


Figure 4.4 Plot of the MAE vs. the latent factors for the COL data on another neural network training.

Table 4.3 Summary of the NLPLS results on the COL data.

NLPLS	R-Sq	R-Sq- Adj	MSE	RMSE	MAE	E-mod	CN	Normwt	N
4 Lat. Factors (A)	0.9942	0.9942	36.8616	6.0714	3.472	0.9457		1	4
2 Lat Factor (B)	0.9958	0.9958	26.7676	5.1737	3.385	0.9471		1	2
5 Lat. Factors (C)	0.9964	0.9964	22.7552	4.7702	3.4819	0.946		1	5

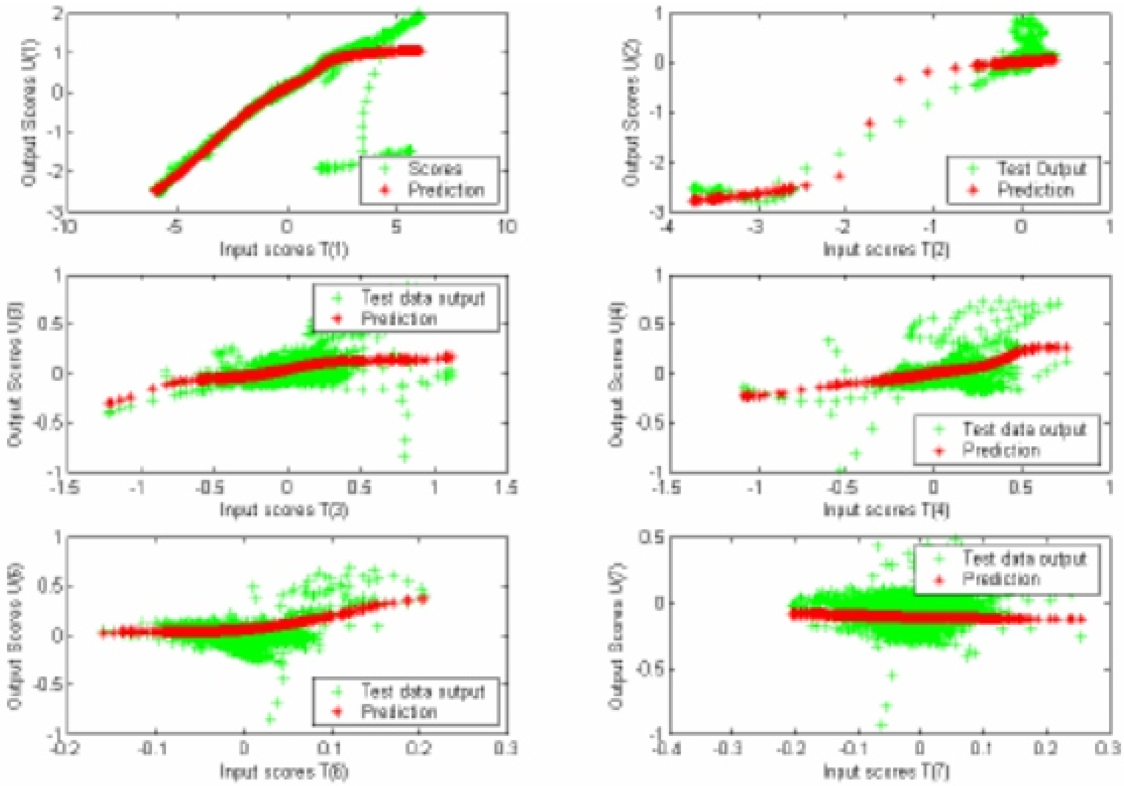


Figure 4.5 Output scores over the input scores (predicted and test response).

4.3 SIMULATED DATA SET ANALYSIS

The Simulated data set was introduced in this section . This data set has a total of 44 variables. The response variable is the 38th variable (this occupied the 38th column before data preprocessing).

4.4 Partial Least Squares on Simulated Data Set

Four models were built with PLS on the simulated data set. Using Malinowski's reduced eigenvalues plot (Figure 4.8), the minimum reduced Eigen factor is not obvious. Factor numbers 3 and 5 were used to build models and the results are given in table 4.4 (first two results in the table). Then the iterative method (generalization) was used to find the optimal number of factors to get a minimum MSE. Figure 4.9 was used to find the optimal number of factors. Looking at the plot, at point 8 latent factors, the line touched the x-axis and ran parallel to that axis. Hence factors 8 and 43 were used to build the model. The summary of the PLS results is shown in Table 4.4. From Table 4.4, the best solution using PLS was the model built with the optimal number of factors (8) from the iterative (generalization) method. It has the best MSE compared to the others, and the condition number was below 100. Figure 4.10 shows the plot of the internal scores vs. the predicted internal scores.

GENERAL RESULTS AND CONCLUSION:-

This chapter gives a conclusion about the various predictive data-mining techniques discussed in this Research paper. It also gives some recommendations for future research in the area of predictive data-mining techniques, both for the linear and the nonlinear models. In selecting the best model in the whole group, five measuring criteria were considered over the nine used in Chapter Four. Models with condition numbers below 100 were chosen first . Then, those with the lowest MSE among those that passed the condition number were chosen. If there were ties, the MAE was used to break the tie; models with lower MAE were chosen over others. If there was still a tie, the modified coefficient of efficiency was used, models with higher modified coefficient of efficiency were favored over others at this stage. Finally, the number of variables, factors or PCs that made the model was used to select the best model. Models with fewer variables, factors or PCs were favored over others. MAE is especially useful in resolving the problem MSE has with outliers. MSE is not regarded as a better measuring criterion than MAE.

SUMMARY OF THE RESULTS OF PDM TECHNIQUES:-

This section gives the summary results of the various techniques used in this Research paper work and how they performed in each type of data set. It also covers the advantages of each technique over the other.

COL Data Results Summary for All the Techniques:-

The COL is a highly ill-conditioned data set. The best model for this data set is the Partial Least Squares (PLS), built with two factors. Among all the models that are below the condition number of 100, it has the best MSE and MAE. The second best is the PCR model with only two PCs. It has MSE lower than the group that survived the condition number elimination. Again, the NLPLS gave the best model if MSE is the only criterion for comparison. It can be seen that the solution is not stable. It has three optimal solutions. Each time the model was retrained, a new optimal solution is obtained. It mapped nonlinearity into the model and hence can only be useful if a nonlinear model is being considered.

Simulated Data Results Summary for All the Techniques:-

The Simulated data set with many input variables has the PLS with optimal factors as the best model for prediction. This is followed by the PCR model with 29 PCs. Both models gave the same measurements, but the PLS came up better by the number of factors used in the model. It used only 8 out of 43 factors for its prediction. The NLPLS also mapped nonlinearity into the model. This is not good because the relationship of the regression is a linear one.

CONCLUSION

In conclusion, in the course of this work, the various data preprocessing techniques were used to process the two data sets. Some of the data sets were seen to have unique features; an example is the COL data set, which is very collinear. PLS generally performed better than all the other four techniques in building linear models. It dealt with the co linearity in the COL data and gave the simplest model that made the best predictions. The PLS also reduced the dimensionality of the data. The study shows that supervised techniques demonstrated a better predictive ability than unsupervised techniques. It can be seen that in MLR and PCR, the correlation-based models which were supervised techniques performed reasonably better than most models where variables and PCs were randomly selected to build the model. The variables that added valuable information to the prediction models were variables that had correlation with the output being predicted. PLS with two factors. From the analysis, it can be seen that the condition number of any data matrix has a direct relation to the number of statistically significant variables in the model. Based on the results from the Summary Tables and the discussion so far on the predictive linear modeling techniques.

RECOMMENDATIONS FOR FUTURE WORK:-

Some predictive data-mining problems are of the non-linear type. For very complex prediction (or forecasting) problems, non-linear algorithms or a blend of both linear and non-linear will be best. This means blends of different algorithms or techniques which combine strengths will be more useful. Effort should be geared towards building super models that combines two or more of these techniques. A great deal of works is going on in evaluating the strengths of the techniques: the neural network (NN), support vector regression (SVR), the regression trees, kernel regression, kernel SVR and so on. Some of the breakthroughs are the kernel support vector machines, the kernel PCA and the least square support vector machines. Another area of data-mining that has been and will continue to be a fertile ground for researchers is the area of data acquisition and storage. The ability to correctly acquire, clean, and store data sets for subsequent mining is no small task. A lot of work is going on in this area to improve on what is obtainable today. There are many commercial software packages produced to solve some of these problems but most are uniquely made to solve particular types of problem. It would be desirable to have mining tools that can switch to multiple techniques and support multiple outcomes. Current data-mining tools operate on structured data but most of the data in the field are unstructured. Since large amount of data are acquired, for example in the World Wide Web, there should be tools that would manage and mine data from this source, a tool that can handle dynamic data, sparse data, incomplete or uncertain data. The dream looks very tall but given the time and energy invested in this field and the results which are produced, it will not be long to get to the development of such software.

REFERENCES:-

1. Lyman, P., and Hal R. Varian, "How much storage is enough?" Storage, 1:4 (2003).

2. Way, Jay, and E. A. Smith,"Evolution of Synthetic Aperture Radar Systems and Their Progression to the EOS SAR," IEEE Trans. Geoscience and Remote Sensing, 29:6 (1991), pp. 962-985.

3. Usama, M. Fayyad, "Data-Mining and Knowledge Discovery: Making Sense Out of Data," Microsoft Research IEEE Expert, 11:5. (1996), pp. 20-25.

4. Berson, A., K. Thearling, and J. Stephen, Building Data Mining Applications for CRM, USA, McGraw-Hill (1999).

5. Malinowski, E. R., "Determination of the Number of Factors and The Experimental Error in a Data Matrix." Anal. Chem. 49 (1977), pp. 612-617.

6. Sharma, K. Sanjay, et al., "A Covariance-Based Nonlinear Partial Least Squares Algorithm," Intelligent Systems and Control Research Group (2004).

7. Giudici, P., Applied Data-Mining: Statistical Methods for Business and Industry. West Sussex, England: John Wiley and Sons (2003).

8. Berry, M. J. A., and G. S. Linoff, Mastering Data Mining. New York: Wiley (2000).

9. Han, J., and M. Kamber, Data Mining: Concepts and Techniques. New York: Morgan Kaufman (2000).

10. Pregibon, D., "Data Mining," Statistical Computing and Graphics, pp. 7-8. (1997).

11. Usama, M. Fayyad, et al., Advances in Knowledge Discovery and Data Mining. Cambridge, Mass.: MIT Press (1996).

12. Ralph, Kimball, The Data Warehouse Toolkit: Practical Technique for Building Dimensional Data Warehouses. New York: John Wiley (1996).

ISSN (Print):2321–404X, (Online):2321–4384

Using of Process Models in Software Engineering

Manoj Kumar
Department of Computer Science
Govt. College, Narnaul

INTRODUCTION:

Software engineering goes through a series of passages that account for their inception, initial development, productive operation, upkeep, and retirement from one generation to another. This article categorizes and examines a number of methods for describing or modelling how software systems are developed. It begins with background and definitions of traditional software life cycle models that dominate most textbook discussions and current software development practices. This is followed by a more comprehensive review of the alternative models of software evolution that are of current use as the basis for organizing software engineering projects and technologies.

Keywords: Software, Process, System, development.

Background: These classic software life cycle models usually include some version or subset of the following activities:

System Planning: New feasible systems replace or supplement existing information processing mechanisms whether they were previously automated, manual.

Requirement Analysis: Identifies the problems a new software system is suppose to solve, its operational capabilities, its desired performance characteristics, and the resource infrastructure needed to support system operation and maintenance.

Prototyping: Identifies and potentially formalizes the objects of computation, their attributes and relationships, the operations that transform these objects, the constraints that restrict system behavior, and so forth.

Architectural Design: Defines the interconnection and resource interfaces between system subsystems, components, and modules in ways suitable for their detailed design and overall configuration management.

Component Implementation and Debugging: Codifies the preceding specifications into operational source code implementations and validates their basic operation.

Software Integration and Testing: Affirms and sustains the overall integrity of the software system architectural configuration through verifying the consistency and completeness of implemented modules, verifying the resource interfaces and interconnections against their specifications, and validating the performance of the system and subsystems against their requirements.

Documentation and Delivery: packaging and rationalizing recorded system development descriptions into systematic documents and user guides, all in a form suitable for dissemination and system support. Deployment and Installation: providing directions for installing the delivered software into the local computing environment, configuring operating systems parameters and user access privileges, and running diagnostic test cases to assure the viability of basic system operation.

Software Maintenance: sustaining the useful operation of a system in its target environment by providing requested functional enhancements, repairs, performance.Software life cycle model: A prescriptive model prescribes how a new software system should be developed. Prescriptive models are used as guidelines or frameworks to organize and structure how software development activities should be performed, and in what order. Typically, it is easier and more common to articulate a prescriptive life cycle model for how software systems should be developed. This is possible since most such models are intuitive or well-reasoned. This means that many idiosyncratic details that describe how a software system is built in practice can be ignored, generalized, or deferred for later consideration. This, of course, should raise concern for the relative validity and robustness of such life cycle models when developing different kinds of application systems, in different kinds of development settings, using different programming languages, with differentially skilled staff, etc. However, prescriptive models are also used to package the development tasks and techniques for using a given set of software engineering tools or environment during a development project.

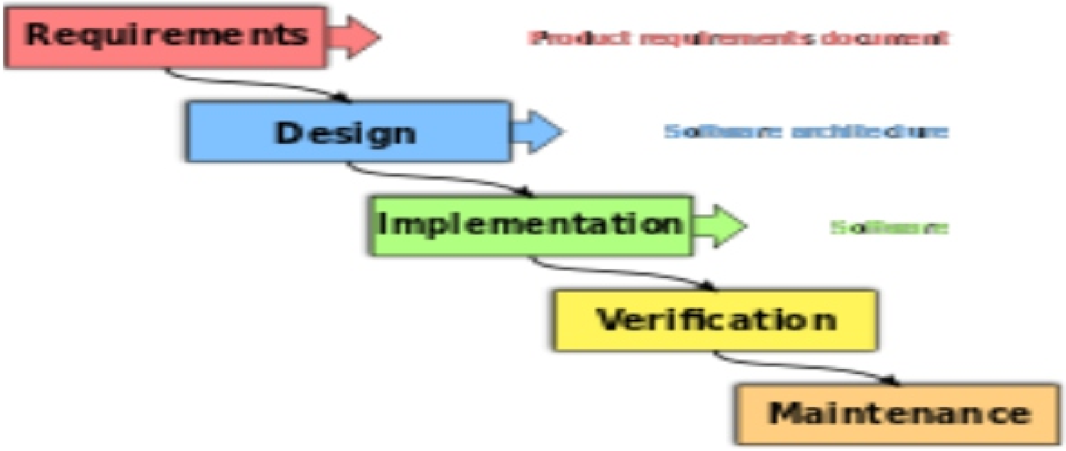


Fig. Software development life cycle or process

Software process model: Software process models often represent a networked sequence of activities, objects, transformations, and events that embody strategies for accomplishing software evolution. Such models can be used to develop more precise and formalized descriptions of software life cycle activities. Their power emerges from their utilization of a sufficiently rich notation, syntax, or semantics, often suitable for computational processing. Software process networks can be viewed as representing multiple interconnected task chains (Kling 1982, Garg 1989). Task chains represent a non-linear sequence of actions that structure and transform available computational objects (resources) into intermediate or finished products. Non-linearity implies that the sequence of actions may be non-deterministic, iterative, accommodate multiple/parallel alternatives, as well as partially ordered to account for incremental progress. Task actions in turn can be viewed a non-linear sequences of primitive actions which denote atomic units of computing work, such as a user's selection of a command or menu entry using a mouse or keyboard. Winograd and others have referred to these units of cooperative work between people and computers as "structured discourses of work" (Winograd 1986), while task chains have become popularized under the name of "workflow" (Bolcer 1998).

Software Production Process Models: There are two kinds of software production process models: non- operational and operational. Both are software process models. The difference between the two primarily stems from the fact that the operational models can be viewed as computational scripts or programs: programs that implement a particular regimen of software engineering and development. Non-operational models on the other hand denote conceptual approaches that have not yet been sufficiently articulated in a form suitable for codification or automated processing.

Non-Operational Process Models: There are two classes of non-operational software process models of the great interest. These are the spiral model and the continuous transformation models. There is also a wide selection of other non-operational models, which for brevity we label as miscellaneous models. Each is examined in turn.

CONCLUSIONS:

Modelling these patterns can utilize features of traditional software life cycle models, as well as those of automatable software process models. Nonetheless, we must also recognize that the death of the traditional system life cycle model may be at hand. New models for software development enabled by the Internet, shifting business imperatives in response to these conditions are giving rise to a new generation of software processes and process models. These new models provide a view of software development and evolution that is incremental, iterative, on-going, interactive, and sensitive to social and organizational circumstances, while at the same time, increasingly amenable to automated support, facilitation, and collaboration over the distances of space and time.

REFERENCES:

1. *Boehm, B., Software Engineering, IEEE Trans. Computer, C-25,12,1226-1241, 1976. Boehm, B. W., Software Engineering Economics, Prentice-Hall, Englewood Cliffs, N. J., 1981*
2. *Basili, V. R., and A. J. Turner, Iterative Enhancement: A Practical Technique for Software Development, IEEE Trans. Software Engineering, 1,4, 390-396, 1975.*
3. *Batory, D., V. Singhal, J. Thomas, S. Dasari, B. Geraci, M. Sirkin, The GenVoca model of software-system generators, IEEE Software, 11(5), 89-94, September 1994.*
4. *Bauer, F. L., Programming as an Evolutionary Process, Proc. 2nd. Intern. Conf. Software Engineering, IEEE Computer Society, 223-234, January, 1976.*
5. *Beck, K. Extreme Programming Explained, Addison-Wesley, Palo Alto, CA, 1999.*
6. *Bendifallah, S., and W. Scacchi, Understanding Software Maintenance Work, IEEE Trans. Software Engineering, 13,3, 311-323, 1987.*
7. *Bendifallah, S. and W. Scacchi, Work Structures and Shifts: An Empirical Analysis of Software Specification Teamwork, Proc. 11th. Intern. Conf. Software Engineering, IEEE Computer Society, 260- 270, 1989.*
8. *Biggerstaff, T., and A. Perlis (eds.), Special Issues on Software Reusability, IEEE Trans. Software Engineering, 10, ,5, 1984.*
9. *Boehm, B., A Spiral Model of Software Development and Enhancement, Computer, 20(9), 61-72, 1987. Boehm, B., A. Egyed, J. Kwan, D. Port, A. Shah, and R. Madachy, Using the WinWin Spiral Model: A Case Study, Computer, 31(7), 33-44, 1998.*
10. *Bolcer, G.A., R.N. Taylor, Advanced workflow management technologies, Software Process-- Improvement and Practice, 4,3, 125-171, 1998.*
11. *Budde, R., K. Kuhlenkamp, L. Mathiassen, and H. Zullighoven, Approaches to Prototyping, Springer- Verlag, New York, 1984.*
12. *Chatters, B.W., M.M. Lehman, J.F. Ramil, and P. Werwick, Modeling a Software Evolution Process: A Long-Term Case Study, Software Process-Improvement and Practice, 5(2-3), 91-102, 2000.*
13. *Cook, J.E., and A.Wolf, Discovering models of software processes from event-based data, ACM Trans. Softw. Eng. Methodol. 7, 3 (Jul. 1998), 215 - 249*
14. *B. Curtis, H. Krasner, V. Shen, and N. Iscoe, On Building Software Process Models Under the Lamppost, Proc. 9th. Intern. Conf. Software Engineering, IEEE Computer Society, Monterey, CA, 96- 103, 1987.*

15. *Curtis, B., H. Krasner, and N. Iscoe, A Field Study of the Software Design Process for Large Systems, Communications ACM, 31, 11, 1268-1287, November, 1988.*
16. *Cusumano, M. and D. Yoffie, Software Development on Internet Time, Computer, 32(10), 60-69, 1999. Distaso, J., Software Management--A Survey of Practice in 1980, Proceedings IEEE, 68,9,1103-1119, 1980.*
17. *DiBona, C., S. Ockman and M. Stone, Open Sources: Voices from the Open Source Revolution, O'Reilly Press, Sebastopol, CA, 1999.*

ISSN (Print):2321–404X, (Online):2321–4384

Security in Wireless Sensor Networks: With Respect to the Latest Issues and Challenges

Meena Chaudhary[1], Dr. Rajeev Kumar[2]
Department of Computer Science and Engineering
[1,2]Sri Venkateshwara University, Gajaraula (Amroha), U.P. India

ABSTRACT

Wireless Sensor Network (WSN) is a developing innovation that shows extraordinary guarantee for different cutting edge applications both for mass open and military. The detecting innovation joined with handling force and remote correspondence makes it lucrative for being abused in wealth in future. The consideration of remote correspondence innovation additionally brings about different sorts of security dangers. The expectation of this paper is to examine the security related issues and difficulties in remote sensor systems. We recognize the security dangers, audit proposed security instruments for remote sensor systems. We likewise talk about the comprehensive perspective of security for guaranteeing layered and powerful security in remote sensor systems.

Keywords Sensor, Security, Attack, Holistic, Challenge.

1. INTRODUCTION

Wireless Sensor Networks (WSN) are developing as both a vital new level in the IT environment and a rich area of dynamic examination including equipment and framework outline, organizing, circulated calculations, programming models, information administration, security and social elements [1]. The fundamental thought of sensor system is to scatter small detecting gadgets; which are fit for detecting a few changes of episodes/parameters and speaking with different gadgets, over a particular geographic territory for some particular purposes like target following, observation, ecological checking and so forth. Today's sensors can screen temperature, weight, dampness, soil cosmetics, vehicular development, clamor levels, lighting conditions, the nearness or nonappearance of specific sorts of items or substances, mechanical anxiety levels on joined articles, and different properties [2]. If there should arise an occurrence of remote sensor organize, the correspondence among the sensors is done utilizing remote handsets. The appealing components of the remote sensor systems pulled in numerous scientists to chip away at different issues identified with these sorts of systems Security is an extensively utilized term enveloping the qualities of confirmation, respectability, protection, nonrepudiation, and hostile to playback [3]. The more the reliance on the data gave by the systems has been expanded, the more the danger of secure transmission of data over the systems has expanded.

2. SECURITY OF DATA IN WSN

2.1 Cryptography

The encryption-unscrambling systems contrived for the conventional wired systems are not achievable to be connected specifically for the remote systems and specifically for remote sensor systems. WSNs comprise of small sensors which truly experience the ill effects of the absence of preparing, memory and battery power [4]. Applying any encryption plan requires transmission of additional bits, consequently additional preparing, and memory and battery power which are vital assets for the sensors' life span. Applying the security instruments, for example, encryption could likewise build deferral, jitter and bundle misfortune in remote sensor systems [5]. Additionally, some basic inquiries emerge while applying encryption plans to WSNs like, how the keys are produced or dispersed. How the keys are overseen, repudiated, appointed to another sensor added to the system or restored for guaranteeing vigorous security for the Network.

2.2. Steganography

While cryptography aims at hiding the content of a message, steganography [6] aims at hiding the existence of the message. Steganography is the art of covert communication by embedding a message into the multimedia data (image, sound, video, etc.) [7]. The main objective of steganography is to modify the carrier in a way that is not perceptible and hence, it looks just like ordinary. It hides the existence of the covert channel, and furthermore, in the case that we want to send a secret data without sender information or when we want to distribute secret data publicly, it is very useful. However, securing wireless sensor networks is not directly related to steganography and processing multimedia data (like audio, video) with the inadequate resources [8] of the sensors is difficult and an open research issue.

2.3 Physical Layer Secure Access

Vital focuses in physical layer secure access are the effective plan so that the bouncing succession is adjusted in less time than is required to find it and for utilizing this both the sender and collector ought to keep up a synchronized clock. A plan as proposed in [9] could likewise be used which presents secure physical layer access utilizing the solitary vectors with the channel blended adjustment.

3. SECURITY THREATS AND ISSUES IN WIRELESS SENSOR NETWORKS

The majority of the dangers and assaults against security in remote systems are practically like their wired partners while some are exacerbated with the consideration of remote network. Truth be told, remote systems are typically more helpless against different security dangers as the unguided transmission medium is more powerless to security assaults than those of the guided transmission medium. The communicate way of the remote correspondence is a basic contender for listening in. In the majority of the cases different security issues and dangers identified with those we consider for remote specially appointed systems are likewise pertinent for remote sensor systems. These issues are well - listed in some past examines [10] furthermore various security plans are now been proposed to battle against them. Be that as it may, the security systems formulated for remote specially appointed systems couldn't be connected straightforwardly for remote sensor systems on account of the architectural disparity of the two networks. While ad hoc networks are self- organizing, dynamic topology, peer to peer networks formed by a collection of mobile nodes and the centralized entity is absent [11]; the wireless sensor networks could have a command node or a base station (centralized entity, sometimes termed as sink).

The architectural aspect of wireless sensor network could make the employment of a security schemes little bit easier as the base stations or the centralized entities could be used extensively in this case. Nevertheless, the major challenge is induced by the constraint of resources of the tiny sensors. In many cases, sensors are expected to be deployed arbitrarily in the enemy territory (especially in military reconnaissance scenario) or over dangerous or hazardous areas. Therefore, even if the base station (sink) resides in the friendly or safe area, the sensor nodes need to be protected from being compromised.

3.1. Attacks in Wireless Sensor Networks

Attacks against wireless sensor networks could be broadly considered from two different levels of views. One is the attack against the security mechanisms and another is against the basic mechanisms (like routing mechanisms). Here we point out the major attacks in wireless sensor networks.

3.1.1 Denial of Service

Denial of Service (DoS) [12] is created by the accidental disappointment of hubs or malevolent activity. The most straightforward DoS assault tries to debilitate the assets accessible to the casualty hub, by sending additional superfluous bundles and in this manner keeps authentic system clients from getting to administrations or assets to which they are entitled. DoS assault is implied not just for the foe's endeavor to subvert, disturb, or devastate asystem, additionally for any occasion that lessens a system's ability to give an administration. In remote sensor organizes, a few sorts of DoS assaults in various layers may be performed. At physical layer the DoS assaults could stick and altering, at connection layer, crash, fatigue, injustice, at system layer, disregard and insatiability, homing, confusion, dark gaps and at transport layer this assault could be performed by vindictive flooding and desynchronization. The systems to forestall DoS assaults incorporate installment for system assets, pushback, solid confirmation and recognizable proof of activity.

3.1.2 Attacks on Information in transit

In a sensor network, sensors monitor the changes of specific parameters or values and report to the sink according to the requirement. While sending the report, the information in transit may be altered, spoofed, replayed again or vanished. As wireless communication is vulnerable to eavesdropping, any attacker can monitor the traffic flow and get into action to interrupt, intercept, modify or fabricate [13] packets thus, provide wrong information to the base stations or sinks. As sensor nodes typically have short range of transmission and scarce resource, an attacker with high processing power and larger communication range could attack several sensors at the same time to modify the actual information during transmission.

3.1.3 Sybil Attack

By and large, the sensors in a remote sensor system may need to cooperate to achieve an errand, thus they can utilize dispersion of subtasks and excess of data. In such a circumstance, a hub can put on a show to be more than one hub utilizing the personalities of other true blue hubs (Figure 1). This kind of assault where a hub produces the characters of more than one hub is the Sybil assault. Sybil assault tries to corrupt the honesty of information, security and asset usage that the circulated calculation endeavors to accomplish. Sybil assault can be performed for assaulting the dispersed stockpiling, steering system, information total, voting, reasonable asset distribution and bad conduct discovery. Fundamentally, any shared system (particularly remote specially appointed systems) is helpless against sybil assault. Be that as it may, as WSNs can have some kind of base stations or passages, this assault could be anticipated utilizing productive conventions. Douceur demonstrated that, without an intelligently concentrated power, sybil assaults are constantly conceivable aside from under extraordinary and farfetched presumptions of asset equality and coordination among substances. Be that as it may, discovery of sybil hubs in a system is not all that simple. The radio asset testing to recognize the nearness of sybil node(s) in sensor arrange and demonstrated that the likelihood to identify the presence of a sybil hub is:

$$Pr(detection) = 1 - (1 - \sum_{all S, M, G} \frac{\binom{s}{S}\binom{m}{M}\binom{g}{G}}{\binom{n}{c}} \frac{S-(m-M)}{c})^{r}$$

Where, n is the number of nodes in a neighbor set, s is the number of sybil nodes, m malicious nodes, g number of good nodes, c is the number of nodes that can be tested at a time by a node, of which S are sybil nodes, M are malicious (faulty) nodes, G are good (correct) nodes and r is the number of rounds to iterate the test.

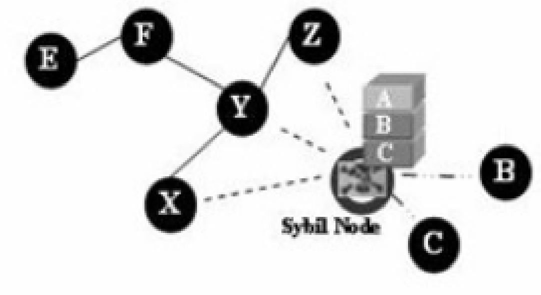


Figure 1: Sybil Attack

Device has been able to insert itself between the communicating nodes (for example, sink and sensor node), it is able to do anything with the packets passing between them. In fact, this attack can affect even the nodes those are considerably far from the base stations. Figure 2 shows the conceptual view of a blackhole/sinkhole attack.

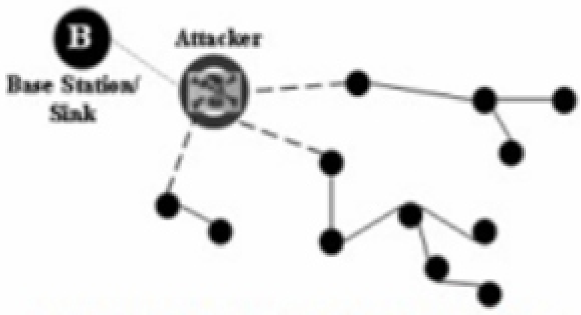


Figure 2: Conceptual view of Blackhole Attack

Hello Flood Attack

Hello Flood Attack is introduced in. This attack uses HELLO packets as a weapon to convince the sensors in WSN. In this sort of attack an attacker with a high radio transmission (termed as a laptop- class attacker in) range and processing power sends HELLO packets to a number of sensor nodes which are dispersed in a large area within a WSN. The sensors are thus persuaded that the adversary is their neighbor. As a consequence, while sending the information to the base station, the victim nodes try to go through the attacker as they know that it is their neighbor and are ultimately spoofed by the attacker.

3.1.7 Wormhole Attack

Wormhole attack is a critical attack in which the attacker records the packets (or bits) at one location in the network and tunnels those to another location. The tunneling or retransmitting of bits could be done selectively. Wormhole attack is a significant threat to wireless sensor networks, because; this sort of attack does not require compromising a sensor in the network rather, it could be performed even at the initial phase when the sensors start to discover the neighboring information.

3.1.4 Blackhole/Sinkhole Attack

In this attack, a malicious node acts as a blackhole to attract all the traffic in the sensor network. Especially in a flooding based protocol, the attacker listens to requests for routes then replies to the target nodes that it contains the high quality or shortest path to the base station. Once the malicious Figure 3 (a and b) shows a situation where a wormhole attack takes place.

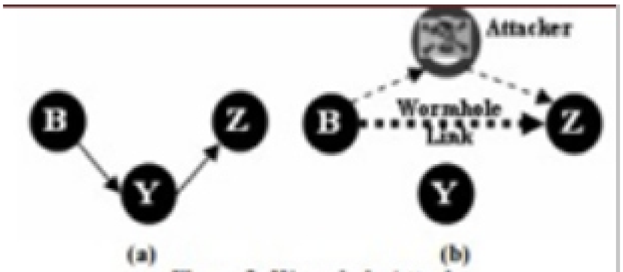


Figure 3: Wormhole Attack

When a node B (for example, the base station or any other sensor) broadcasts the routing request packet, the attacker receives this packet and replays it in its neighborhood. Each neighboring node receiving this replayed packet will consider itself to be in the range of Node B, and will mark this node as its parent. Hence, even if the victim nodes are multi- hop apart from B, attacker in this case convinces them that B is only a single hop away from them, thus creates a wormhole.

4. PROPOSED SECURITY SCHEMES AND RELATED WORK

In the recent years, wireless sensor network security has been able to attract the attentions of a number of researchers around the world. In this section we review and map various security schemes proposed or implemented so far for wireless sensor networks.

4.1. Security Schemes for Wireless Sensor Networks

It gives an analysis of secure routing in wireless sensor networks. Studies how to design secure distributed sensor networks with multiple supply voltages to reduce the energy consumption on computation and therefore to extend the network's life time. aims at increasing energy efficiency for key management in wireless sensor networks and uses Younis et. al. network model for its application. Wood et al. studies DoS attacks against different layers of sensor protocol stack. JAM presents a mapping protocol which detects a jammed region in the sensor network and helps to avoid the faulty region to continue routing within the network, thus handles DoS attacks caused by jamming.

In the maker's exhibit that wormholes those are so far considered terrible for WSN could enough be used as a responsive resistance instrument for thwarting staying DoS ambushes? Ye et. al. Presents a truthful in transit filtering (SEF) instrument to recognize injected false data in sensor framework and focus principally on the most ideal approach to channel false data using total secret and thusly keeping any single exchanged off center point from breaking the entire system. SNEP and µTESLA are two secure building ruins for giving data mystery, data freshness and convey approval. Tiny Sec proposes an association layer security instrument for sensor frameworks which uses a profitable symmetric key encryption tradition.

Table 1: Summary of various security schemes for wireless sensor networks

Security Schemes	Attacks Deterred	Network Architecture	Major Features
JAM	DoS Attack (Jamming)	Traditional wireless sensor network	Avoidance of jammed region by using coalesced neighbor nodes
Wormhole based	DoS Attack (Jamming)	Hybrid (mainly wireless partly wired) sensor network	Uses wormholes to avoid jamming
Statistical En-Route Filtering	Information Spoofing	Large number of sensors, highly dense wireless sensor network	Detects and drops false reports during forwarding process

Radio Resource Testing, Random Key Pre-distribution etc.	Sybil Attack	Traditional wireless sensor network	Uses radio resource, Random key pre-distribution, Registration procedure, Position verification and Code attestation for detecting sybil entity
Bidirectional Verification, Multi-path multi-base station routing	Hello Flood Attack	Traditional wireless sensor network	Adopts probabilistic secret sharing, Uses bidirectional verification and multi-path multi- base station routing
On Communication Security	Information or Data Spoofing	Traditional wireless sensor network	Efficient resource management, Protects the network even if part of the network is compromised
TIK	Wormhole Attack, Information or Data Spoofing	Traditional wireless sensor network	Based on symmetric cryptography, Requires accurate time synchronization between all communicating parties, implements temporal leashes
Random Key Predistribution	Data and information spoofing, Attacks in information in Transit Data and Information	Traditional wireless sensor network Distributed Sensor Network, Large- scale wireless sensor network with Spoofing	Provide resilience of the network, Protect the network even if part of the network is compromised, Provide authentication measures for sensor nodes Suitable for large wireless sensor networks which allows addition dynamic nature node capture
REWARD	Blackhole attacks	Traditional wireless sensor network	Uses geographic routing, Takes advantage of the broadcast inter-radio behavior to watch neighbor transmissions and detect
TinySec	Data and Information spoofing, Message Replay Attack	Traditional wireless sensor network	blackhole attacks Focuses on providing message authenticity, integrity and
SNEP & µTESLA	Data and Information Spoofing, Message Replay Attacks	Traditional wireless sensor network	Semantic security, Data authentication, Replay protection, Weak freshness, Low communication overhead

Proposes separate security plans for information with different affectability levels and an area - based plan for remote sensor organizes that ensures whatever is left of the system, notwithstanding when parts of the system are traded off executes symmetric key Cryptographic calculations with postponed key divulgence on bits to build up secure correspondence channels between a base station and sensors inside its extent.

4.2. Holistic Security in Wireless Sensor Networks

A holistic approach aims at improving the performance of wireless sensor networks with respect to security, longevity and connectivity under changing environmental conditions.

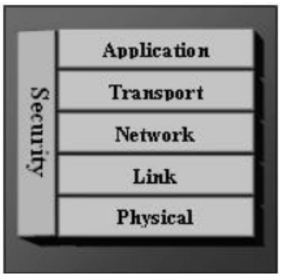


Figure 4: Holistic view of Security in wireless sensor networks

The holistic approach has some basic principles like, in a given network; security is to be ensured for all the layers of the protocol stack, the cost for ensuring security should not surpass the assessed security risk at a specific time, if there is no physical security ensured for the sensors, the security measures must be able to exhibit a graceful degradation if some of the sensors in the network are compromised, out of order or captured by the enemy and the security measures should be developed to work in a decentralized fashion.

5. CONCLUSION

The vast majority of the assaults against security in remote sensor systems are brought on by the inclusion of false data by the traded off hubs inside the system. For guarding the consideration of false reports by bargained hubs, a method is required for distinguishing false reports. Be that as it may, growing such a discovery component and making it proficient speaks to an incredible examination challenge. Once more, guaranteeing all encompassing security in remote sensor system is a noteworthy exploration issue. A hefty portion of todays proposed security plans depend on particular system models. As there is an absence of joined push to take a typical model to guarantee security for every layer, in future however the security components turn out to be entrenched for every individual layer, consolidating every one of the instruments together to make them work in a joint effort with each other will acquire a hard research challenge. Regardless of the fact that all encompassing security could be guaranteed for remote sensor organizes, the cost-adequacy and vitality effectiveness to utilize such components could at present stance incredible exploration challenge in the coming days.

REFERENCES

[1] Culler, D. E and Hong, W., "Wireless Sensor Networks", Communication of the ACM, Vol. 47, No. 6, June 2004, pp. 30-33.

[2] Pathan, A-S. K., Islam, H. K., Sayeed, S. A., Ahmed, F. and Hong, C. S., "A Framework for Providing E-Services to the Rural Areas using Wireless Ad Hoc and Sensor Networks", to appear in IEEE ICNEWS 2006.

[3] Undercoffer, J., Avancha, S., Joshi, A., and Pinkston, J., "Security for Sensor Networks", CADIP Research Symposium, 2002, available at, http://www.cs.sfu.ca/~angiez/personal/paper/ sensor-ids.pdf

[4] Perrig, A., Szewczyk, R., Wen, V., Culler, D., and Tygar, J. D., "SPINS: Security Protocols for Sensor Networks", Wireless Networks, vol. 8, no. 5, 2002, pp. 521-534.

[5] Saleh, M. and Khatib, I. A., "Throughput Analysis of WEP Security in Ad Hoc Sensor Networks", Proc. The Second International Conference on Innovations in Information Technology (IIT'05), September 26-28, Dubai, 2005.

[6] Kurak, C and McHugh, J, "A Cautionary Note on Image Downgrading in Computer Security Applications", Proceedings of the 8th Computer Security Applications Conference, San Antonio, December, 1992, pp. 153-159.

[7] Kim, C. H., O, S. C., Lee, S., Yang, W. I., and Lee, H-W., "Steganalysis on BPCS Steganography", Pacific Rim Workshop on Digital Steganography (STEG'03), July 3-4, Japan , 2003.

[8] Younis, M., Akkaya, K., Eltoweissy, M., and Wadaa, A., "On handling QoS traffic in wireless sensor networks", Proc. of the 37th Annual Hawaii International Conference on System Sciences, 2004, 5-8 January, 2004, pp. 292-301.

[9] Orihashi, M., Nakagawa, Y., Murakami, Y., and Kobayashi, K., "Channel synthesized modulation employing singular vector for secured access on physical layer", IEEE GLOBECOM 2003, Volume 3, 1-5 December, 2003, pp. 1226 – 1230.

[10] Zhou, L. and Haas, Z. J., "Securing ad hoc networks", IEEE Network, Volume 13, Issue 6, ov.-Dec. 1999, pp. 24 – 30. Strulo, B., Farr, J., and Smith, A., "Securing Mobile Ad hoc Networks — A Motivational Approach", BT Technology Journal, Volume 21, Issue 3, 2003,
pp. 81 – 89.

[11] Wang, B-T. and Schulzrinne, H., "An IP traceback mechanism for reflective DoS attacks", Canadian Conference on Electrical and Computer Engineering, Volume 2, 2-5 May 2004, pp. 901 – 904.

[12] Douceur, J. "The Sybil Attack", 1st International Workshop on Peer-to-Peer Systems (2002).

[13] Newsome, J., Shi, E., Song, D, and Perrig, A, "The sybil attack in sensor networks: analysis & defenses", Proc. of the third international symposium on Information processing in sensor networks, ACM, 2004, pp. 259-268.

ISSN (Print):2321–404X, (Online):2321–4384

An Assessment of Data Transmission in Wireless Sensor Networks with Enhancement to the Security and Reliability

Anu Chaudhary[1], Dr. Rajeev Kumar[2]
Department of Computer Science and Engineering
[1,2]Sri Venkateshwara University, Gajaraula (Amroha), U.P. India

ABSTRACT

Wireless sensor networks can be used in a wide assortment of uses running from combat zone observation in military, through remote patient checking in pharmaceutical to woodland fire identification in natural applications. Lion's share of WSN applications requires at any rate some level of security. With a specific end goal to accomplish the required level, secure and hearty directing is vital. Secure information transmission is a basic issue for wireless sensor networks (WSNs). Bunching is a powerful and down to earth approach to improve the framework execution of WSNs. Bunch based information transmission in WSNs has been explored by researchers to accomplish the network adaptability and administration, which boosts hub lifetime and decrease data transfer capacity utilization by utilizing neighborhood joint effort among sensor hubs. In this work, we have outlined a directing convention named solid and secure information transmission convention (RSDT) which is secure and dependable and the outcomes will be contrasted and other steering conventions of same class, for example, secure and proficient information transmission (SET) conventions for WSNs, called SET-IBOOS, the personality based online/disconnected computerized signature (IBOOS) plan. RSDT will utilize same idea of signature as utilized as a part of SET-IBOOS.

Keywords: WSN, Secure WSN, Energy Efficient WSN, Hierarchical routing, cluster head

1. INTRODUCTION

Wireless sensor networks comprise of numerous little conservative gadgets, furnished with sensors (e.g. acoustic, seismic or picture sensors), that frame a wireless network. Every sensor hub in the network gathers data from its environment, and sends it to a base station, either from sensor hub to sensor hub i.e. multi jump, or specifically to a base station i.e. single jump [1].A wireless sensor network may comprise of hundreds or up to a large number of sensor hubs and can be spread out as a mass or set out one by one. The sensor hubs work together with each other over a wireless media to build up a detecting network, i.e. a wireless sensor network. Due to the conceivably vast size of the wireless sensor networks, every individual sensor hub must be little and of ease. The accessibility of minimal effort sensor hubs has brought about the advancement of numerous other potential application territories, e.g. to screen vast or unfriendly fields, woods, houses, lakes, seas, and procedures in enterprises. The sensor network can give access to data by gathering, handling, examining and conveying information from nature [2]. In numerous application ranges the wireless sensor network must have the capacity to work for drawn out stretches of time, and the unwavering quality and also security of transmitting information is vital [1].

2. RELATED WORK

In this paper [1], The creators propose two secure and proficient information transmission (SET) conventions for CWSNs, called SET-IBS and SET- IBOOS, by utilizing the character based computerized signature (IBS) plan and the personality based online/disconnected advanced mark (IBOOS) plan, individually. In SET-IBS, security depends on the hardness of the Diffie-Hellman issue in the matching space. SET-IBOOS further diminishes the computational overhead for convention security, which is vital for WSNs, while its security depends on the hardness of the discrete logarithm issue. The creators demonstrate the attainability of the SET-IBS and SET-IBOOS conventions concerning the security necessities and security investigation against different assaults. In [2], the creators detail the mystery sharing-based multipath routing issue as animprovement issue. The disjoint multipath routing plan with mystery sharing is generally perceived as one of the successful routing techniques to guarantee the wellbeing of data. This sort of plan changes every parcel into a few shares to improve the security of transmission. A three-stage disjoint routing plan called the Security and Energy-proficient Disjoint Route (SEDR) is proposed. In light of the mystery sharing calculation, the SEDR plot dispersively and arbitrarily conveys shares everywhere throughout the network in the initial two stages and afterward transmits these shares to the sink hub. The creators in [3] displayed another routing instrument, which incorporates FEC codes and particular encryption plan for giving both QoS and secure information transmission in WSN. In the proposed convention, RS coding is utilized to give dependability and security. The sink hub settles on the ways determination process keeping in mind the end goal to fulfill the unwavering quality or the deferral prerequisites by an application and the quantity of these ways is resolved to improve the dependability. The exploration article [4] presents the advancement of the genuine proving ground of BIOSARP routing convention. The proving ground comprises of 10 sensor hubs (TELOSB). The BIOSARP routing convention based WSN performs well and its versatile conduct to the natural changes guarantees solid information exchange.

3. IBOOS SCHEME FOR CLUSTERED WIRELESS SENSOR NETWORK

An IBOOS plan [1] executed for CWSNs comprises of taking after four operations, particularly, setup at Each node senses the environment at a fixed rate and always has data to send to the base station.

- All sensor nodes are immobile.
- The Key concept of SET-IBOOS is used as it eases. So, we assume that we are securing the data in same way as in IBOOS.

$$
\begin{aligned}
T(n) &= \frac{\rho}{1-\rho \times \left(r \bmod \left\lfloor \frac{1}{\rho} \right\rfloor\right)} \cdot \frac{E_{cur}(n)}{E_{init}(n)} && \forall n \in G_n, \\
T(n) &= 0 && \forall n \notin G_n.
\end{aligned}
$$

the BS, key extraction and disconnected marking at the CHs, web marking at the information sending hubs, and check at the accepting hubs [1]:

I. Setup. The BS (as a trust power) produces an expert key msk and open parameters param for the private key generator (PKG), and offers them to all sensor hubs.

II. Extraction. Given an ID string, a sensor hub produces a private key sekID connected with the ID utilizing msk.

III. Offline marking. Given open parameters and timestamp t, the CHsensor hub creates a disconnected mark SIG disconnected, and transmits it to the leaf hubs in its cluster.

IV. Online marking. From the private key sekID, SIG disconnected and message M, a sending hub (leaf hub) produces an online mark SIG on the web.

V. Verification. Given ID, M, and SIG on the web, the accepting hub (CH hub) online is substantial, and output

In Step 1, toward the start of the setup period of another round, the BS first communicates its ID, a nonce (number utilized once), and the meaning of the beginning time Ts of the current round to all sensor hubs, which is utilized for the mark marking and confirmation in the setup stage. In Step 2, a sensor hub chooses whether to wind up a CH for the current round, taking into account the limit T(n) contrasted with numbers from 0 with 1, which is set as takes after:

Dynamic clustering calculation ideally with duplicating the proportion of leftover vitality of the present sensor hub (i.e., Ecur(n)/Einit(n)) to expand the vitality proficiency in the clustering, where Ecur(n) is the present vitality, and Einit(n) is the underlying vitality of the sensor hub. ρ is from the earlier decided worth which remains for the wanted rate of CHs amid one round (e.g., ρ = 10%), r is the current round number, and Gn is the arrangement of sensor hubs that have not been CHs in the last 1/ρ rounds. On the off chance that the estimation of decided number is not exactly the limit, the sensor hub chooses itself as a CH. The sensor hub who chooses to end up a CH communicates the notice message (adv) to the neighboring hubs in the network, which is linked with the mark. In Step 3, the sensor hub, which chooses to be a leaf hub, picks a CH to join in view of the biggest got signal quality of adv messages. At that point, it speaks with CH I by sending a join_request (join) message, which is linked with the goal CH's ID IDi, its own ID IDj, time stamp Ts, and the computerized signature

4. METHODOLOGY

The establishment of proposed convention lies in the acknowledgment that the base station is a high-vitality hub with a lot of vitality supply. In this manner, proposed convention uses the base station to control the planned detecting undertaking performed by the sensor hubs. In proposed convention, the accompanying suspicion is to be considered.

- A settled base station is situated in focal point of the locale
- The sensor hubs are vitality compelled with a uniform introductory vitality assignment.
- The hubs are furnished with force control capacities to shift their transmitted force.
- Each hub detects the earth at an altered rate and dependably has information to send to the base station.
- All sensor hubs are fixed.
- The Key idea of SET-IBOOS is utilized as it simplicity. Along these lines, we accept that we are securing the information in same path as in IBOOS.

The radio channel should be symmetrical. In this manner, the vitality required to transmit a message from a source hub to a goal hub is the same as the vitality required to transmit the same message from the goal hub back to the source hub for a given SNR (Signal to Noise Ratio). In addition, it is accepted that the correspondence environment is dispute and blunder free. Thus, there is no requirement for retransmission.

The means of the calculation are:

1. At first, base station is set up at the focal point of district and hubs are setup in that specific area and every hub will have level with vitality.
2. In cycle 1, Cluster Head will be made by condition.
3. The choice of every hub to end up cluster head is taken taking into account the recommended rate of cluster head hubs p. A sensor hub picks an arbitrary number, r, somewhere around 0 and 1. On the off chance that this arbitrary number is not exactly a limit esteem, T (n), the hub turns into a cluster-head for the current round. The limit worth is ascertained in view of a condition that fuses the wanted rate to wind up a cluster-head, the current round, and the arrangement of hubs that have not been chosen as a cluster-head in the last (1/P) rounds, signified by G. T (n) is given in eq. (1)

Ideal number of cluster heads is assessed to be 10% of the aggregate number of hubs.

Likelihood of hub to wind up cluster head will fluctuate as indicated by given condition:

$$p\ (I)= P * n * RE *EN / (TE* Ea) \qquad (2)$$

where,
P= probability to become cluster head which is 0.1.
n = no. of nodes
RE= remaining energy of a particular node

EN = Initial energy of a particular node
TE = Total Energy of the Network
Ea = Average energy of network Average energy of network is given by equation:
Average energy= TE * (1-r / rmax) / n; (3)

where,
TE = Total Energy of the Network
r = current round
rmax= total no. of rounds
n = no. of nodes

Now, the modified formula for choosing cluster head is given in eq (1)

4. At that point, Nodes sends the information to their particular cluster heads and vitality utilization will be computed.
5. Cluster Head will total the information and send it to the base station and vitality utilization will be computed for every hub and cluster heads
6. In cycle 2, the hubs will get to be cluster heads as per likelihood condition i.e. as per least separation from base station and limit vitality.
7. After choice of cluster heads, Nodes sends the information to their individual cluster heads, that will be chosen by least separation of a specific hub from cluster heads and vitality utilization will be ascertained.
8. Cluster Head will total the information and send it to the base station and vitality utilization will be figured.
9. This procedure will be rehashed until the entire network gets down or number of rounds wrapped up.
10. Execution will be assessed by like network lifetime, vitality dissemination, no. of information bundles sent and so forth

5. IMPLEMENTATION AND RESULTS

5.1 Parameter Value

Table 1.Network Parameters

Network field:	100 x100m
N (Number of nodes):	100
Initial energy:	1 J
Eelec (ETx&ERx):	50nJ/bit

ε fs (free:	2 10space)pJ/bit/m
εmp:	0.0013 pJ/bit/m^4
E$_{DA}$:	5 nJ/bit/signal
Eoff:	5 μJ/ signature
Eon:	12.37 μJ/ signature
Data packet size:	4000 bits
Tool used:	MATLAB 7.6.0

Figure 1 shows the deployment of nodes and base station in a particular region. The region we have taken for simulation is 100m x 100m. The 'o' symbol denotes the base station (sink). The position of nodes is taken similar in both techniques IBOOS as well as in RSDT protocol.

Figure 1: Deployment of nodes and base station

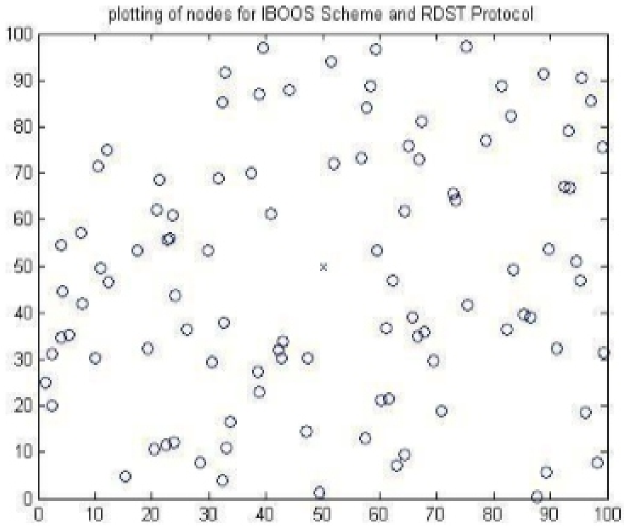


Figure 2: Number of Rounds vs Number of Nodes Dead

Figure 2 demonstrates the correlation of routing conventions character based online/disconnected computerized signature (IBOOS) plan, and dependable and secure information transmission convention (RSDT) regarding Number of hubs dead. Figure 2 demonstrates the general lifetime of the network. Here, we can watch that RSDT performs better in contrast with IBOOS in sending information to base station. Around 80% of network performs better in contrast with IBOOS.

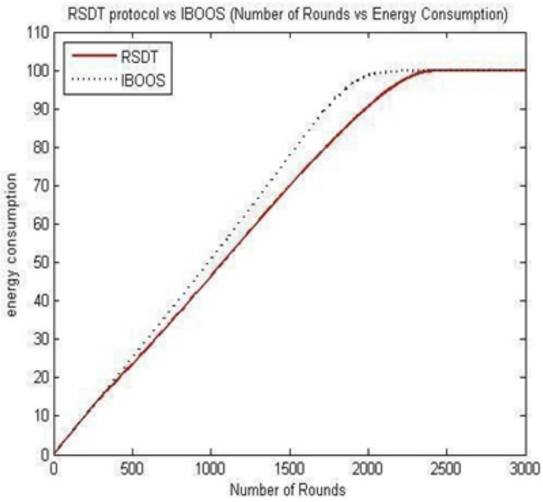


Figure 3: Number of Rounds vs Energy Consumption

Figure 3 shows the lifetime of the network. It shows that how energy of the network consumes step by step and finally whole network goes down. It can be observed from the figure denotes the nodes and 'x' 3 that, RSDT consumes less energy and sustain more number of rounds as compare to IBOOS protocol.

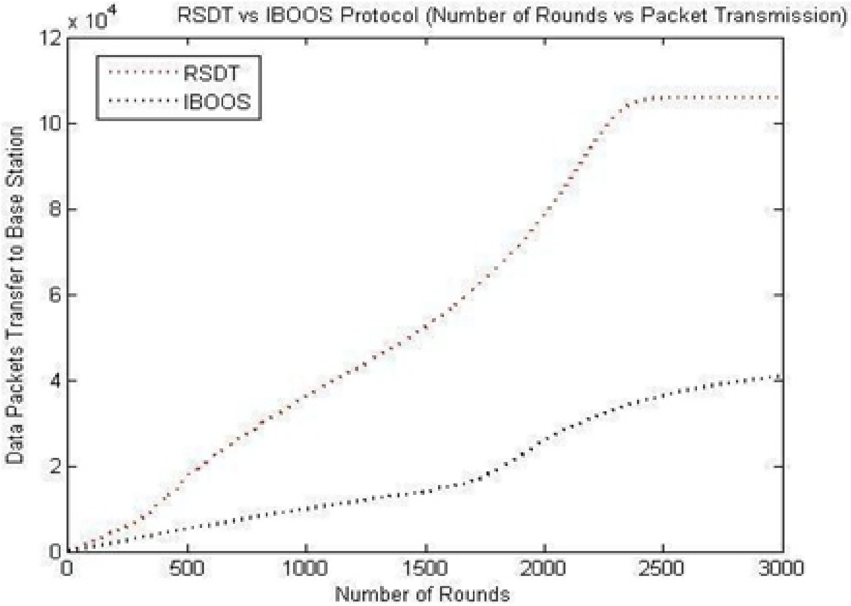


Figure 4: Number of Rounds vs Data Packets sent to base station

Figure 4 shows how much data will be sent from nodes to SINK. From figure 4, we can observed that, in IBOOS protocol data sent to base station is relatively less as compared to RSDT.

Table 2: Network Lifetime

When First Node dies		When 30% of Nodes dies		When 80% of Nodes dies	
IBOOS	RSDT	IBOOS	RSDT	IBOOS	RSDT
1713	2070	1930	2274	2299	2383

Table 2 shows the network lifetime comparison between IBOOS and RSDT protocol. IBOOS first node is died after 1713 rounds whereas RSDT first node died after 2070 rounds. More survival of rounds means network lifetime will increases, which finally increases the packet transfer to base station and hence RSDT will become more reliable as compare to IBOOS.

6. CONCLUSION AND FUTURE SCOPE

WSNs vary from conventional wireless correspondence networks in a few of their qualities like dependability of information, force mindfulness, security and so forth. This new routing convention named Reliable and secure information transmission convention (RSDT) which is progressive routing based with the entire control to the base station or we can say that base helped. While self-designing, the hubs are unconscious about the entire coherent structure of the network. Be that as it may, in Reliable and secure information transmission convention (RSDT), the base station first gathers data about the consistent structure of the network and lingering vitality of every hub. In this way, with the worldwide data about the network base station improves as in it has data about the lingering vitality of every hub. At last, Reliable and secure information transmission convention (RSDT) is contrasted and right now created routing convention IBOOS. A correlation of these is done on the premise of vitality scattering with time, framework lifetime of network and number of bundles sent to base station. In WSN, hundreds or a great many sensor hubs are haphazardly scattered in the sensor field. These hubs sense the information and send this detected information to the cluster head (if there should be an occurrence of various leveled routing) or specifically to the base station as per the TDMA (time division multiplexing access) given by cluster head or base station separately. In future, the work can be upgrade to plan a routing convention which is more vitality proficient and secure for Wireless Sensor Networks by applying hereditary calculation or some other improvement system to settle base station in a region where it can cover most extreme hubs.

REFERENCES

[1] Huang Lu, "Secure and Effi for Cluster-Based Wireless Sensor TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 3, MARCH 2014 pp. 750- 761.

[2] Anfeng Liu, ZhongmingZheng, "Secure Energy- Efficient Disjoint Multipath Routing for WSNs",IEEE TRANSACTIONS ON VEHICULARTECHNOLOGY, VOL. 61, NO. 7, SEPTEMBER2012.

[3] Hind Alwan, and Anjali Ag MECHANISM FOR QOS ROUTING IN WIRELESS SENSOR NIEEET WORKS",2012.

[4] KashifSaleem-Time Empirical,"AStudyRealof BIOSARP based Wireless Sensor Network Testb ed",IEEE2012.

[5] Lynda Mokdad, "Performan security routing strategies to avoid DoS attacks in WSN",IEEE, 2012.

[6] Sankardas Roy, "Securenin D Wireless Sensor Networks: Filtering out the Attacker's IEEEImpact TRANSACTIONS", ONINFORMATION FORENSICS AND SECURITY,VOL. 9, NO. 4, APRIL 2014.

[7] KashifSaleem, "Empirical-inspired Self- Organized Secure Autonomous Routing ProtocolIEEE,2013",.

[8] WenjunGu, NeelanjanaDutta, Sriram Chellappan, and XiaoleBai,-to - End"ProvidinSecure Communications in WirelesIEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 8, NO. 3,SEPTEMBER 2011.

[9] Huei-Wen Ferng and Dian Rachmarini , Routing Protocol for Wireless Sensor Networks with Consideration ofIEEEnergy,2012. Ef

[10] SushmitaRuj, AmiyaNayak, Key Distribution in Wireless Sensor Networks with Applications",IEEETR ANSACTIONS ONCOMPUTERS, VOL. 62, NO. 11, NOVEMBER2013.

[11] A. SelcukUluagac,-BAsed "SecurLoose Synchronization (SOBAS) for Wireless Sensor Networks",IEEET RANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 4,APRIL 2013.

[12] HimaliSaxena, "DSF-A Distributed Security Framework for Heterogeneous Wireless Sensor Networks", IEEE2010.

[13] Leron Lightfoot, -Location"Preserving Source Privacy in Wireless Sensor Network using STaR Routing",I EEE2010.

ISSN (Print):2321–404X, (Online):2321–4384

Study of Wireless Lan Technology: Special Reference to Ieee 802.11 Standards

Gaurav Kumar[1], Dr. Senthilkumar A[2]

Department of Electronics and Communication Engineering

[1,2],Himalayan University, Arunachal Pradesh (India)

ABSTRACT

Arrange advancements are generally in light of wireline arrangements. Be that as it may, the presentation of the IEEE 802.11 principles have had a colossal effect available with the end goal that portable workstations, PCs, printers, cellphones, and VoIP telephones, MP3 players in our homes, in workplaces and even out in the open territories have consolidated the remote LAN innovation. Remote broadband advancements these days give boundless broadband access to clients which were beforehand offered just to wireline clients. In this paper, we audit and compress one of the rising remote broadband innovation i.e. IEEE 802.11,which is an arrangement of physical layer standard for executing remote neighborhood PC correspondence in the 2.4,3.6,5 and 60GHz recurrence band. They settle innovation issues or add usefulness which is relied upon to be required by future applications. Despite the fact that a portion of the prior renditions of these innovations are out of date, (for example, HiperLAN) now yet we have included them in this audit for culmination.

Keywords: Wireless Communications, IEEE802.11, HiperLAN, WLAN, Wi-fi.

1. INTRODUCTION

The remote broadband innovations were produced with the point of giving administrations similar to those gave to the wire line systems. Cell arranges now offer help for high transmission capacity information exchange for various portable clients all the while. What's more, they likewise give portability support to voice correspondence. Remote information systems can be isolated into a few sorts relying upon their zone of scope. They are:

WLAN: Wireless Local Area network, in Zone with a cell range up to hundred meters, basically in home and office conditions [1].

WMAN: Wireless Metropolitan Area Network; for the most part cover more extensive regions as substantial as whole urban communities. WWAN: Remote Wide Area Network with a cell run around 50 km, cover domains greater than a city [2].

However out of these norms, WLAN and late advancements in WLAN innovation would be our principle range of study in this paper. The IEEE 802.11 is the most broadly sent WLAN innovation starting today. Another outstanding is the HiperLAN standard by ETSI. Both these advancements are joined under the Wireless Fidelity (Wi-fi) organization together. In writing however, IEEE802.11 and Wi-fi is utilized conversely and we will likewise take after a similar tradition in this paper. A normal WLAN arrange comprises of an Access Point (AP) in the center/focus and various stations (STAs) are associated with this focal Access Point (AP).Now, there are fundamentally two modes in which correspondence can happen [3].

2. DEVELOPMENT OF IEEE 802.11

The Physical layer (PHY) and medium get to control (MAC) layer were mostly focused by the IEEE 802 venture. Right when the likelihood of remote neighborhood (WLAN) was at first envisioned, it was as of late pondered another PHY of one of the available rules. The principal competitor which was considered for this was IEEE's most conspicuous standard 802.3. However later discoveries demonstrated that the radio medium carried on very unique in relation to the customary all around acted wire. As there was lessening indeed, even over short detachments, impacts couldn't be perceived. Henceforth, 802.3's bearer sense various access with impact location (CSMA/CD) couldn't be connected [4].

The following applicant standard considered was 802.4. By then of time, its planned medium get to i.e. the token transport idea was accepted to be better than 802.3's conflict based plan. Henceforth, WLAN started as 802.4L. Later in 1990 it got to be distinctly evident that token taking care of in radio systems was fairly troublesome. The institutionalization body understood the need of a remote correspondence standard that would have its own exceptionally one of a kind MAC. At long last, on March 21, 1991, the venture 802.11 was affirmed (figure 1).

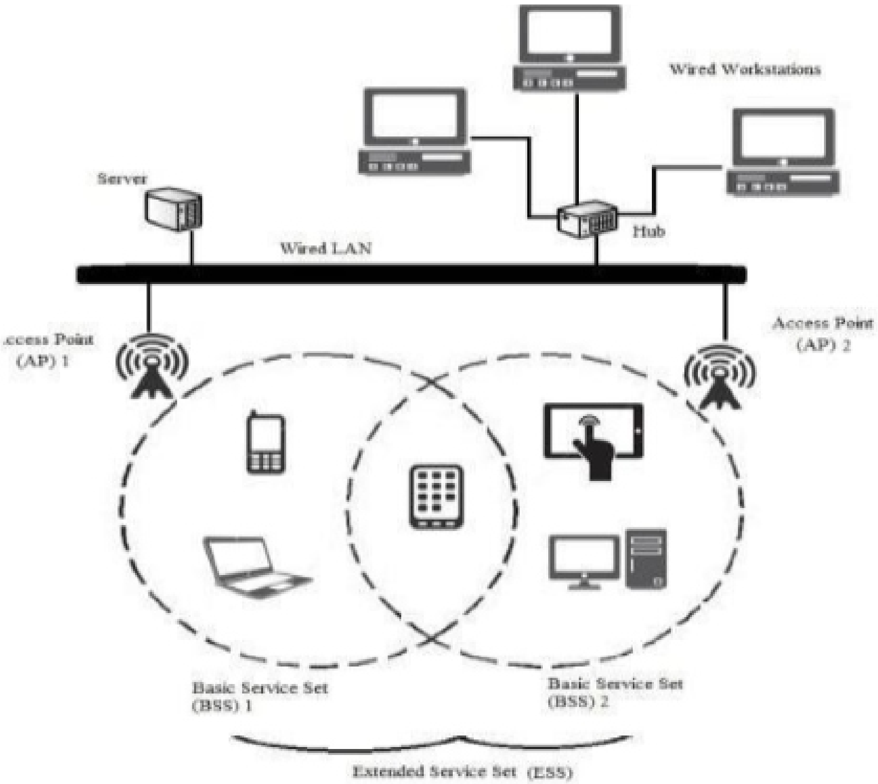

Wired Workstations
Server
Wired LAN
Hub
.ccess Point (AP) 1
Access Point (AP) 2
Basic Service Set (BSS) 1
Basic Service Set (BSS) 2
Extended Service Set (ESS)

Figure 1 WLAN Network Architecture

3. IEEE 802.11 FAMILY

The most generally sent 802.11 standard has a considerable measure of augmentation and numerous more are at present a work in progress. Initially presented in 1999,the IEEE 802.11 gauges was basically created remembering the home and the workplace condition for remote neighborhood. The Initial measures gave a most extreme information rate of 2Mbps for every AP which expanded to 11 Mbps per AP with the arrangement of IEEE 802.11b.Newer expansions like IEEE 802.11g and IEEE 802.11a gave greatest information rate of 54Mbps for every AP utilizing different strategies to help up the most extreme information rates [5]. WLAN gadgets in light of IEEE 802.11g at present offer information rate 100-125Mbps [6].

3.1. Physical (PHY) Layer

The IEEE 802.11 utilizations assortment of PHY layers with the point of expanding the total throughput of the system. IEEE 802.11 standard incorporates three PHY layers in particular:

1. FHSS (Frequency Hopping Spread Spectrum)
2. Table 1OFDM PHY layer modulation techniques

Data Rate (Mbps)	Modulation	Coding rate	Coded bits/sub Carrier	Code bits/OFDM symbol	Data bits/OFDM symbol
6	BPSK	01-Feb	1	48	24
9	BPSK	03-Apr	1	48	36
12	QPSK	01-Feb	2	96	48
18	QPSK	03-Apr	2	96	72
24	16-QAM	01-Feb	4	192	96
36	16-QAM	03-Apr	4	192	144
48	64-QAM	02-Mar	6	288	192
54	64-QAM	03-Apr	6	288	216

4. CONCLUSION

A portion of the reasons which can be referred to for such across the board utilization of WLANs are low framework cost, simplicity of advancement, support for versatile client correspondence, organization without cabling and simplicity of adding new client to the system bringing about a tremendous decline in execution cost. As the significance of portable client has expanded complex, WLANs have increased much significance in homes, schools, workplaces and so on and developed as a get to innovation in short separation interchanges. Still today, WLANs experiences a considerable measure of issues. A standout amongst the most vital disadvantages is the utilization of shared medium in which execution gets extensively corrupted as the quantity of STAs increments in the WLAN arrange. The issue of unapproved get to and spying in WLANs are a portion of the genuine security issues which have been a long standing cerebral pain for the IEEE working gathering. Diverse security encryption plans had been actualized previously. Be that as it may, so far all such encryption frameworks have been demonstrated to have security vulnerabilities.

REFERENCES

1. *IEEE 802.11-1999, IEEE Standard for Local and Metropolitan Area Networks Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, June 12, 1999.*
2. *IEEE 802.11b-1999, IEEE Standard for Local and Metropolitan Area Networks Specific Requirements – Part 11: Wireless LAN Medium*
3. *Access Control (MAC) and Physical Layer (PHY) Specifications High Speed Physical Layer Extension in the 2.4 GHz Band, September 16, 1999*
4. *IEEE 802.11a-1999, IEEE Standard for Local and Metropolitan Area Networks Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications High Speed Physical Layer in the 5 GHz Band, 1999.*
5. *IEEE 802.11g-2003, IEEE Standard for Local and Metropolitan Area Networks Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band, June 27, 2003.*
6. *U.S. Apply autonomy, 802.11g Speed Acceleration How We Do It, www.usr.com/download/whitepap ers/1 25mbps-wp.pdf 2004.*

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005