SPECIAL ISSUE

THE IIOAB JOURNAL

VOLUME 10 : NO 2 : JULY 2019 : ISSN 0976-3104

Institute of Integrative Omics and
Applied Biotechnology Journal

Dear Esteemed Readers, Authors, and Colleagues,

I hope this letter finds you in good health and high spirits. It is my distinct pleasure to address you as the Editor-in-Chief of Integrative Omics and Applied Biotechnology (IIOAB) Journal, a multidisciplinary scientific journal that has always placed a profound emphasis on nurturing the involvement of young scientists and championing the significance of an interdisciplinary approach.

At Integrative Omics and Applied Biotechnology (IIOAB) Journal, we firmly believe in the transformative power of science and innovation, and we recognize that it is the vigor and enthusiasm of young minds that often drive the most groundbreaking discoveries. We actively encourage students, early-career researchers, and scientists to submit their work and engage in meaningful discourse within the pages of our journal. We take pride in providing a platform for these emerging researchers to share their novel ideas and findings with the broader scientific community.

In today's rapidly evolving scientific landscape, it is increasingly evident that the challenges we face require a collaborative and interdisciplinary approach. The most complex problems demand a diverse set of perspectives and expertise. Integrative Omics and Applied Biotechnology (IIOAB) Journal has consistently promoted and celebrated this multidisciplinary ethos. We believe that by crossing traditional disciplinary boundaries, we can unlock new avenues for discovery, innovation, and progress. This philosophy has been at the heart of our journal's mission, and we remain dedicated to publishing research that exemplifies the power of interdisciplinary collaboration.

Our journal continues to serve as a hub for knowledge exchange, providing a platform for researchers from various fields to come together and share their insights, experiences, and research outcomes. The collaborative spirit within our community is truly inspiring, and I am immensely proud of the role that IIOAB journal plays in fostering such partnerships.

As we move forward, I encourage each and every one of you to continue supporting our mission. Whether you are a seasoned researcher, a young scientist embarking on your career, or a reader with a thirst for knowledge, your involvement in our journal is invaluable. By working together and embracing interdisciplinary perspectives, we can address the most pressing challenges facing humanity, from climate change and public health to technological advancements and social issues.

I would like to extend my gratitude to our authors, reviewers, editorial board members, and readers for their unwavering support. Your dedication is what makes IIOAB Journal the thriving scientific community it is today. Together, we will continue to explore the frontiers of knowledge and pioneer new approaches to solving the world's most complex problems.

Thank you for being a part of our journey, and for your commitment to advancing science through the pages of IIOAB Journal.

Yours sincerely,

Vasco Azevedo

Vasco Azevedo, Editor-in-Chief
Integrative Omics and Applied Biotechnology
(IIOAB) Journal

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

FACTORS IMPACTING TO SMART CITY IN VIETNAM WITH SMARTPLS 3.0 SOFTWARE APPLICATION

Bui Huy Khoi*, Nguyen Thi Ngan
Faculty of Business Administration, Industrial University of Ho Chi Minh City, VIETNAM

ABSTRACT

In recent times, the domestic and foreign economic sciences have done a lot of research on developing smart cities. However, the scientific argument for the solution remains open, especially the quantitative model of the factor that affect the development of smart cities. Therefore, we collected data for seeking practical evidence for the model, sample size is 314. The content of the article focuses on 2 main issues: theoretical framework of the quantitative model and the results of the application for building smart city model. The reliability and value of the scale are determined by Confirmatory factor analysis (CFA). The results of Structural equation modeling (SEM)by use Smart pls 3.0 Software show that smart city is influenced by 6 factors.

KEY WORDS
Vietnam, Smart city, Pc, Pvc, SEM, Smartpls 3.0

INTRODUCTION

As we know, in the period of industrialization, modernization and social development, the emerging problem is "the development of smart cities" in Vietnam. By the end of 2016, the rate of urbanization nationwide is about 36.7%, equivalent to 33.62 million people, up from 1% in 2015 (equivalent to 1.35 million people). Urbanization and industrialization are always dependent, complementary. The clear evidence for this combination is reflected in the rapidly growing number of industrial parks in Vietnam coupled with the urbanization process. To better understand this issue, the group made this essay to better understand "Smart Urban Development in Vietnam during the Industrial Revolution 4.0".

A smart city must have smart defensive solutions to combat smart modern weapons. We need to learn the experiences of countries that have succeeded in the smart urban model in the world such as Korea, Japan and Singapore, as well as exchanging experiences with other countries (Sweden, Spain, Germany, etc.), and it is important to improve the policy institutions, while at the same time providing appropriate legal documents for management. To strive to put the urban centers of Vietnam into a smart urban center so as to raise the economic efficiency and help the living environment better and people are better served and facilitated. for people involved in urban management and government oversight.

Hypothesis development

Received: 4 Oct 2018
Accepted: 26 Nov 2018
Published: 4 Jan 2019

The concept “smart city” was introduced already in 1994. The smart city is the idea of environmental sustainability as its main aim is reducing greenhouse gas emissions in urban areas through the deployment of innovative technologies. The developing interest in the smart city concept and the needs to solve the challenges related to urbanization lead to several private and public investments in the technology development and deployment [1].

Petrolo, R., Loscri, V., & Mitton, N. (2017) suggested there were six factors affecting smart city: smart governance, smart people, smart economy, smart living and smart environment [2].

The smart city model typically integrates the economic, social and environmental components of the city in a way that sustainably maximizes the efficiency of the city’s primary systems. Elhoseny, H., Elhoseny, M., Riad, A. M., & Hassanien, A. E. (2018) offered nine factors related smart city: smart government, smart living, smart business, smart education, smart mobility, smart utility and smart environment [3].

Talari, S., Shafie-khah, M., Siano, P., Loia, V., Tommasetti, A., & Catalão, J. P. (2017) gave five affected smart city: Smart Communities, Smart Homes, Smart Buildings, Responsive Customers, Smart Energy and Smart Grids.

Kumar, T. V., & Dahiya, B. (2017) emphasize A Smart City System comprises of six key building blocks: (i) smart people, (ii) smart city economy, (iii) smart mobility, (iv) smart environment, (v) smart living, and (vi) smart governance [4].

Smart people

***Corresponding Author**
Email:
buihuykhoi@iuh.edu.vn

There are eleven observations in Smart people: (1) Smart people excel in what they do professionally. (2) Smart people have a high Human Development Index. (3) A smart city integrates its universities and colleges into all aspects of city life. (4) It attracts high human capital, for example knowledge workers. (5) A smart city maintains high Graduate Enrolment Ratio and has people with high level of qualifications and expertise. (6) Its inhabitants opt for lifelong learning and use e-learning models. (7) People in a smart city are highly flexible and resilient to the changing circumstances. (8) Smart city inhabitants excel in creativity and find unique solutions to challenging issues. (9) Smart people are cosmopolitan, are open-minded, and

IIOAB JOURNAL
ISSN: 0976-3104

hold a multicultural perspective. (10) Smart people maintain a healthy lifestyle. (11) Smart people are actively involved in their city's sustainable development, its efficient and smooth functioning, its upkeep and management, and making it more live able [4].

Smart city economy

Smart City Economy has nineteen: (1) A smart city understands its economic DNA. (2) A smart city is driven by innovation and supported by universities that focus on cutting-edge research, not only for science, industry, and business but also for cultural heritage, architecture, planning, development, and the like. (3) A smart city highly values creativity and welcomes new ideas. (4) A smart city has enlightened entrepreneurial leadership. (5) A smart city offers its citizens diverse economic opportunities. (6) A smart city knows that all economics works at the local level. (7) A smart city is prepared for the challenges posed by and opportunities of economic globalization. (8) A smart city experiments, supports, and promotes sharing economy. (9) A smart city thinks locally, acts regionally, and competes globally. (10) A smart city makes strategic investments on its strategic assets. (11) A smart city develops and supports compelling national brand/s. (12) A smart city insists on balanced and sustainable economic development (growth). (13) A smart city is a destination that people want to visit (tourism). (14) A smart city is nationally competitive on selected and significant factors. (15) A smart city is resourceful, making the most of its assets while finding solutions to problems. (16) A smart city excels in productivity. (17) A smart city has high flexibility of labour market. (18) A smart city welcomes human resources that enhance its wealth. (19) A smart city's inhabitants strive for sustainable natural resource management and understand that without this its economy will not function indefinitely [4].

Smart mobility

Smart Mobility includes: (1) A smart city focuses on the mobility of people, and not only that of vehicles. (2) A smart city will advocate walkability and cycling. (3) A smart city has vibrant streets (at no additional cost). (4) A smart city effectively manages vehicular and pedestrian traffic, and traffic congestion. (5) A smart city has pleasurable (bicycle) routes. (6) A smart city has balanced transportation options. (7) A smart city will have mass rapid transit system, such as metro rail, light metro, monorail, or 'skytrain' for high-speed mobility. (8) A smart city will have integrated high-mobility system linking residential areas, work places, recreational areas, and transport notes (e.g. bus/railway station/s and airport). (9) A smart city will practice high-density living, such that benefit of high-speed mobility is uniformly available. (10) A smart city has seamless mobility for differently-abled (often incorrectly called, disabled) people.

Smart environment

Smart Environment is such as: (1) A smart city lives with and protects the nature. (2) A smart city is attractive and has a strong sense of place that is rooted in its natural setting. (3) A smart city values its natural heritage, unique natural resources, biodiversity, and environment. (4) A smart city conserves and preserves the ecological system in the city region. (5) A smart city embraces and sustains biodiversity in the city region. (6) A smart city efficiently and effectively manages its natural resource base. (7) A smart city has recreational opportunities for people of all ages. (8) A smart city is a green city. (9) A smart city is a clean city. (10) A smart city has adequate and accessible public green spaces. (11) A smart city has an outdoor living room. Unlike the indoor living room in houses where we meet others, outdoor living rooms are aesthetically designed intimate, active, and dynamic urban realms where people meet face to face for a culturally and recreationally rich and enjoyable contact as part of living and work. (12) A smart city has distinctive and vibrant neighbor hoods that encourage neighbor liness and a spirit of community. (13) A smart city values and capitalizes on scenic resources without harming the ecological system, natural resources, and biodiversity. (14) A smart city has an integrated system to manage its water resources, water supply system, wastewater, natural drainage, floods and inundation, especially in the watersheds where it is located, especially in view of the (impending) climate change. (15) A smart city focuses on water conservation and minimizes the unnecessary consumption of water for residential, institutional, commercial, and industrial use, especially in the arid and semi-arid areas. (16) A smart city has an efficient management system for the treatment and disposal of wastewater, and reuse of treated wastewater, particularly in the arid and semi-arid areas. (17) A smart city has an efficient management system for the collection, treatment, and disposal of industrial wastewater. (18) A smart city has an integrated and efficient management system for the collection, transfer, transportation, treatment, recycling, reuse, and disposal of municipal, hospital, industrial, and hazardous solid waste. (19) A smart city has an efficient system to control air pollution and maintain clear air, especially in the air sheds where it is located. (20) A smart city has an efficient and effective system for disaster risk reduction, response, recovery, and management. (21) A smart city has and continually upgrades its urban resilience to the impacts of climate change. (22) A smart city can create a low-carbon environment with focus on energy efficiency, renewable energy, and the like.

Smart living

IIOAB JOURNAL
ISSN: 0976-3104

There are fourteen in Smart Living: (1) A smart city has strong and shared values. (2) A smart city records and celebrates local history, culture, and nature. (3) A smart city has a vibrant downtown, 24 h and 7 days a week. (4) A smart city can provide the necessary safety and security to women, children, and senior citizens. (5) A smart city improves the urban way of life. (6) A smart city builds natural and cultural assets to build a good quality of life. (7) A smart city not only understands the big picture of urban liveability, but also pays attention to small details. (8) A smart city has high-quality open and accessible public spaces. (9) A smart city has high-quality public services and amenities. (10) A smart city is an ideal place of living, especially for women, children, and senior citizens. (11) A smart city organizes festivals that celebrate people, life, and nature in city. (12) A smart city has a ritual event (or more) that symbolizes the values and aspirations of the community. (13) A smart city celebrates and promotes art, cultural, and natural heritage in the city. (14) A smart city engages artists to improve and enrich the aesthetics of daily life of the city.

Smart governance

Finally, Smart Governance has (1) A smart city practices accountability, responsiveness, and transparency (ART) in its governance. (2) A smart city uses big data, spatial decision support systems and related geospatial technologies in urban and city regional governance. (3) A smart city constantly innovates e-governance for the benefit of all its residents. (4) A smart city constantly improves its ability to deliver public services efficiently and effectively. (5) A smart city practices participatory policy-making, planning, budgeting, implementation, and monitoring. (6) A smart city has a clear sustainable urban development strategy and perspectives known to all. (7) A smart city utilizes creative urban and regional planning with focus on the integration of economic, social, and environmental dimensions of urban development. (8) A smart city features effective, efficient, and people-friendly urban management. (9) A smart city practices E-Democracy to achieve better development outcomes for all. (10) A smart city embraces a Triple Helix Model in which government, Academia and Business/Industry practice changing roles in Governance.

So, we gave the proposed research hypotheses as following:

“Hypothesis 1 (H1). There is a positive impact of Smart Economy on smart city.
“Hypothesis 2 (H2). There is a positive impact of Smart citizen on smart city.
“Hypothesis 3 (H13. There is a positive impact of Smart governance on smart city.
“Hypothesis 4 (H4). There is a positive impact of Smart life on smart city.
“Hypothesis 5 (H5). There is a positive impact of Smart Environment on smart city.
“Hypothesis 6 (H6). There is a positive impact of Smart Traffic on smart city
All hypotheses and factors are modified as [Fig. 1].

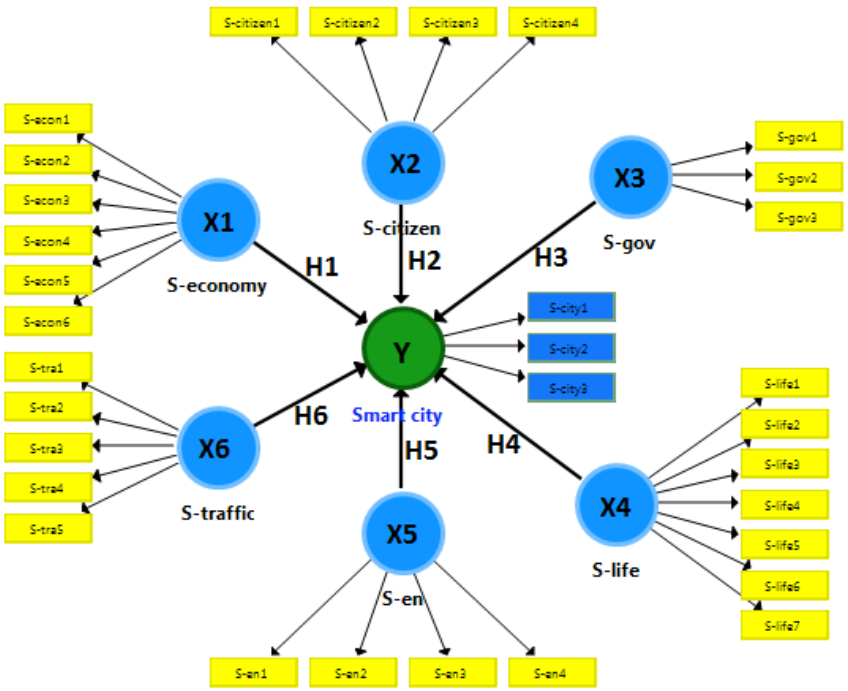

Fig 1: Research model (Source: Designed by author)

.

S-economy (X1): Smart economy

IIOAB JOURNAL
ISSN: 0976-3104

(1) A smart city highly values creativity and welcomes new ideas. (2) A smart city offers its citizens diverse economic opportunities. (3) A smart city is prepared for the challenges posed by and opportunities of economic globalization. (4) A smart city insists on balanced and sustainable economic development (growth). (5) A smart city excels in productivity (6) A smart city's inhabitants strive for sustainable natural resource management and understand that without this its economy will not function indefinitely.

S-citizen (X2): Smart citizen

(1) Smart people excel in what they do professionally. (2) Smart people have a high Human Development Index. (3) Smart city inhabitants excel in creativity and find unique solutions to challenging issues. (4) Smart people are actively involved in their city's sustainable development, its efficient and smooth functioning, its upkeep and management, and making it more liveable

S-gov (X3): Smart governance

(1) A smart city uses big data, spatial decision support systems and related geospatial technologies in urban and city regional governance. (2) A smart city constantly innovates e-governance for the benefit of all its residents. (3) A smart city utilizes creative urban and regional planning with focus on the integration of economic, social, and environmental dimensions of urban development.

S-life (X4): Smart life

(1) A smart city has strong and shared values. (2) A smart city records and celebrates local history, culture, and nature. (3) A smart city can provide the necessary safety and security to women, children, and senior citizens. (4) A smart city builds natural and cultural assets to build a good quality of life. (5) A smart city not only understands the big picture of urban liveability, but also pays attention to small details. (6) A smart city has high-quality open and accessible public spaces. (7) A smart city has high-quality public services and amenities.

S-en (X5): Smart environment

(1) A smart city lives with and protects the nature. (2) A smart city has an efficient management system for the collection, treatment, and disposal of industrial wastewater. (3) A smart city has an efficient and effective system for disaster risk reduction, response, recovery, and management. (4) A smart city can create a low-carbon environment with focus on energy efficiency, renewable energy, and the like.

S-traffic (X6): Smart traffic

(1) A smart city focuses on the mobility of people, and not only that of vehicles. (2) A smart city effectively manages vehicular and pedestrian traffic, and traffic congestion. (3) (4) A smart city will have mass rapid transit system, such as metro rail, light metro, monorail, or 'skytrain' for high-speed mobility. (5) A smart city will practice high-density living, such that benefit of high-speed mobility is uniformly available.

(Y): Smart city,
(1) Smart city is affected by six factors. (2)

METHOD

We followed the methods of Ly H. Anh, Le Si Dong, Vladik Kreinovich, and Nguyen Ngoc Thach (2018) [5]. Research methodology is implemented through two steps: qualitative research and quantitative research. Qualitative research was conducted with a sample of 57 people. Quantitative research was carried out as soon as the question was edited from the test results with a sample of 314 people. According to Hair et al [6], the sample size must be at least ≥ m x 5, in which m is the number of observed variables. So, with 29 variables observed in this study, the sample size should be at least ≥ 145. Therefore, 314 people are surveyed by face-to-face method in Ho Chi Minh City. Respondents were selected by convenient methods with a sample size of 314people in Hochiminh City in Vietnam in [Table 1]. Samples of 52 people was incorrect because they did not have full information in this questionnaire. There were 68 (26%) males and 194 (74%) females in this survey in [Table 1]. The questionnaire answered by respondents is the main tool to collect data.

The survey was conducted in May 2018 in Hochiminh City, Vietnam. Data processing and statistical analysis software is used by Smartpls 3.0 developed by SmartPLS GmbH Company in Germany. The reliability and validity of the scale were tested by Cronbach's Alpha, Average Variance Extracted (Pvc) and Composite Reliability (Pc). Cronbach's alpha coefficient greater than 0.6 would ensure the scale reliability [7].

Composite Reliability (Pc) is better than 0.6 and Average Variance Extracted must be greater than 0.5 [6, 8]. Followed by a linear structural model SEM was used to test the research hypotheses [5, 9].

Table 1: Sample demographic characteristics

Sample		Amount	Percent (%)
SEX	Male	68	26
	Female	194	74
	Total	262	100
QUALIFICATION	Bachelor	228	87
	Master	10	3.8
	Other	24	9.2
	Total	262	100

Source: Calculated by SPSS.sav and Excel.csv

Datasets

We validate our model on three standard datasets for smart city in Vietnam: SPSS. sav, Excel.csv and Smartpls. splsm. Dataset has seven variables: six independent variables and one variable. There are 314 observations and 29 factors in dataset. SPSS. sav and Excel.csv were used for descriptive statistics and Smartpls. splsm for advanced analysis.

Data Availability can receive from author by email.

MEASURES

Structural Equation Modeling (SEM) is used on the theoretical framework. Partial Least Square method can handle many independent variables, even when multi collinearity exists. PLS can be implemented as a regression model, predicting one or more dependent variables from a set of one or more independent variables or it can be implemented as a path model. Partial Least Square (PLS) method can associate with the set of independent variables to multiple dependent variables [5, 9].

Consistency and reliability

In this reflective model convergent validity is tested through composite reliability or Cronbach's alpha. Composite reliability is the measure of reliability since Cronbach's alpha sometimes underestimates the scale reliability [5, 9-11]. [Table 2] shows that composite reliability varies from 0.704 to 0.820 which is above preferred value of 0.5. This proves that model is internally consistent. To check whether the indicators for variables display convergent validity.

Cronbach's alpha is used. From [Table 2], it can be observed that all the factors are reliable (Cronbach's alpha > 0.60 and Pvc > 0.5). Some factors have Cronbach's alpha < 0.60 and Pvc < 0.5 but they have composite reliability (Pc) better than 0.6. So they will be supported and analyze next steps.

Table 2: Cronbach's alpha, composite reliability (Pc) and AVE values (Pvc)

Factor	Cronbach's Alpha	Average Variance Extracted (Pvc)	Composite Reliability (Pc)	*p* Value	Findings
S-citizen	0.584	0.443	0.751	0.000	Supported
S-economy	0.586	0.358	0.750	0.000	Supported
S-en	0.710	0.524	0.814	0.000	Supported
S-gov	0.673	0.602	0.820	0.000	Supported
S-life	0.522	0.268	0.704	0.000	Supported
S-traffic	0.631	0.415	0.771	0.000	Supported

$$P_{Pvc} = \frac{\sum_{i=1}^{p} \gamma_i^2}{\sum_{i=1}^{p} \gamma_i^2 + \sum_{i=1}^{p} (1-\gamma_i^2)}; P_c = \frac{(\sum_{i=1}^{p} \gamma_i)^2}{(\sum_{i=1}^{p} \gamma_i)^2 + \sum_{i=1}^{p} (1-\gamma_i^2)} \qquad \alpha = \frac{k}{k-1}\left[1 - \frac{\sum \sigma^2(x_i)}{\sigma_x^2}\right]$$

Source: Calculated by Smartpls software 3.0.

Structural equation modeling (SEM)

SEM results in the [Fig. 2] showed that the model is compatible with data research. The Smart city is affected by six factors about 73.3%. The six hypotheses are supported because their p-value is greater than 0.05 as [Table 3].

Table 3: Structural Equation Modeling (SEM)

Relation	Beta	SE	T Value	*p* Value	Findings
S-citizen -> Smart city	0.198	0.042	4.739	0.000	Supported
S-economy -> Smart city	0.293	0.053	5.531	0.000	Supported
S-en -> Smart city	0.275	0.038	7.196	0.000	Supported
S-gov -> Smart city	0.218	0.045	4.873	0.000	Supported
S-life -> Smart city	0.306	0.045	6.792	0.000	Supported
S-traffic -> Smart city	0.291	0.043	6.765	0.000	Supported

Beta (r): SE = SQRT(1-r2)/(n-2);CR= (1-r)/SE; P-value =TDIST(CR, n-2, 2).

Source: Calculated by Smartpls software 3.0.

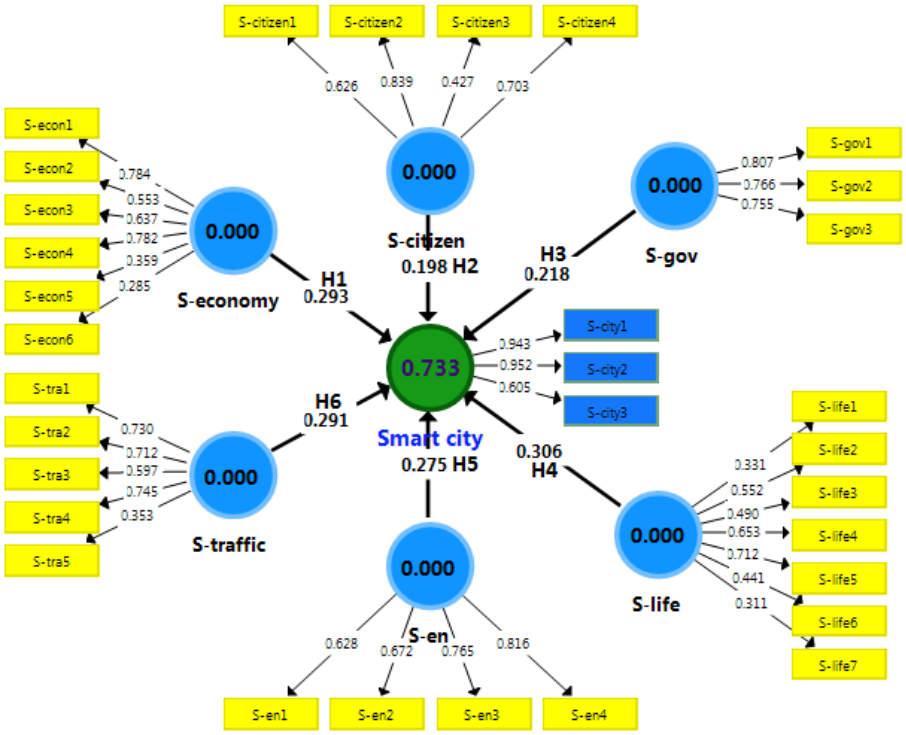

Fig. 2: Structural Equation Modeling (SEM) (Source: Calculated by Smartpls software 3.0.)

SEM results showed that the model is compatible with data research: SRMR has p-value ≤ 0.001 (<0.05) [11, 12]in [Table 4].

Hypotheses H1, H2, H5, H7, H9, H11 and H12 were supported. The results indicated H3, H4, H6, H8, H10 and H13 unsupported.

Table 4: Standard of model SEM

Standard	Beta	SE	*T*-Value	*p*	Findings
SRMR	0.067	0.005	12.674	0.000	Supported

Source: Calculated by Smartpls software 3.0.

In bootstrapping, resampling methods are used to compute the significance of PLS coefficients. Output of significance levels can be retrieved from bootstrapping option. [Table 4] shows the results of hypotheses testing; all the t values above 1.96 are significant at the 0.05 level [5, 9].

DISCUSSION

IIOAB JOURNAL
ISSN: 0976-3104

Advantages

Viet Nam is a little behind compared to the other countries in the world but it has many comparative advantages. Viet Nam has great access to modern technology and we received a lot of help from international experts and organizations. In addition, big corporations like CISCO and IBM and providers of software to manage smart cities have lent the hand in a number of sectors in Viet Nam, such at environment, construction and land management.

Many models are also being applied in some cities such as the Triple Helix Model, connecting government, businesses and scientists. Meanwhile, other models incorporate different sectors. For newly developed cities and localities, the can develop and integrates smart city management from the beginning, which will benefit different groups. This mean that different groups in the city will all contribute to the model and benefit from it.

Disadvantages and challenges

Vietnam is facing many challenges stemming from the urban explosion and rapid urbanization in the country. Therefore, smart urban development is one of the urgent requirements, contributing to control of the environment, infrastructure and social security.

However, planning is slow when there are many challenges in terms of human resources, capital, and lack of a vision of urban planning. Thereby, it can be seen, applying hi-tech system to bring into the urban development concealed many challenges. The development of each urban area is different, the construction and development of Vietnam's construction is not synchronous, the technical infrastructure has not been fully developed. Therefore, when integrating hi-tech systems into urban areas, there will be some difficulties.

- The major difficulties in deploying Smart City in Vietnam include: the cost is too large; The ability to connect information between units is limited; The government has no policies, incentives and human resources. And the biggest barrier to building Smart City is the determination, inertia of the system, psychological hesitation to change due to issues related interests.
- Resources are hindering the Smart City development process. Although TP. Ho Chi Minh City is more convenient than other provinces in terms of economic and human resources, but still have difficulty applying, the order of investment resources can cause waste as technology changes very quickly. In addition, the policy is a barrier, and hope in the future can be resolved to solve problems between the ministries and localities.
- The construction of smart cities is still facing many difficulties due to the technical infrastructure has not kept pace with the development of the city, leading to the consequences such as traffic congestion, lack of water, waste and rain waterlogging, environmental pollution waste.
- Another difficulty, which is the biggest challenge for planning and building smart cities, is a matter of perception. From the needs, awareness of information technology of the people to the awareness of managers. If not aware, look ahead to the development of smart urban and the development of technology will be how, to meet the practical needs in the next years, the quality of life of people is difficult improve. Urban Vietnam is hardly able to keep up with the modern, intelligent development of cities around the world.
- Participation of people is impossible when they have to live in difficult conditions, without technology ... when deploying smart cities.
- Smart city does not require much labor, so many workers lose their jobs, such as traffic police, tourist police. This has led to high unemployment, redundant labor markets, resulting in social evils; Social welfare is burdensome.
- At Smart city, it seems that all basic information of individuals, families and organizations is stored in a center for the purpose of living activities such as security, taxation, employment and residence. The possibility of leakage due to technical errors, as information is sold out, is used in the bad is possible. This information can be used for black business, terrorism, exploitation on the harm each other.

The challenge is to make people believe in Smart City is reasonable, the trend of the times, so that everyone cooperates with each other. In addition, how to find the right direction, when there are resources, there is unity, where should I invest in to press the correct opening of many doors, is the lever of social economy, not wasted resources.

CONCLUSION

The research question was built from the legacy of questions from previous studies, and surveys in developing countries, more specific is Vietnam. The results from the data collected is then analyzed Smartpls 3.0.

- Raising people's awareness to help people come closer to modern and advanced technology.
- Encourage startup projects to attract investment from outside.

- For us, we should be self-conscious of updating new knowledge, cultivating and learning human knowledge.
- The state as well as the managers have to make plans for the city, so that people can trust a smart city in the future. From there, get the support from citizens.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

ACKNOWLEDGEMENTS

This work was supported in part by Industrial University of Ho Chi Minh City, Vietnam.

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Ahvenniemi H, Huovila A, Pinto-Seppä I, Airaksinen M. [2017] What are the differences between sustainable and smart cities. 60:234-245.

[2] Petrolo R, Loscri V, Mitton N. [2017] Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms, Transactions on Emerging Telecommunications Technologies. 28.

[3] Elhoseny H, Elhoseny M, Riad A, Hassanien AE. [2018] A framework for big data analysis in smart cities, in International Conference on Advanced Machine Learning Technologies and Applications. 405-414.

[4] Kumar TV, Dahiya B. [2017] Smart economy in smart cities, in Smart Economy in Smart Cities, ed: Springer. 3-76.

[5] Khoi BH, Van Tuan N. [2018] Using SmartPLS 3.0 to Analyze Internet Service Quality in Vietnam, in Econometrics for Financial Applications, Studies in Computational Intelligence, 760, LH Anh, LS Dong, V Kreinovich, NN Thach, Eds., ed: Springer Nature. 430-439.

[6] Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. [2006] Multivariate data analysis, ed: Upper Saddle River, NJ: Pearson Prentice Hall. Vol. 6. 289.

[7] Nunnally JC, Bernstein I. [1994] The assessment of reliability, Psychometric theory. 3:248-292.

[8] Hair Jr JF, Hult GTM, Ringle C, Sarstedt M. [2016] A primer on partial least squares structural equation modeling (PLS-SEM): Sage Publications. 95-204.

[9] Khoi BH, Van Tuan N. [2018] Using SmartPLS 3.0 to Analyze Internet Service Quality in Vietnam, in International Econometric Conference of Vietnam: Springer. 430-439.

[10] Wong KKK. [2013] Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS, Marketing Bulletin. 24:1-32.

[11] Latan H, Noonan R. [2017] Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications: Springer. Nov 3.

[12] Henseler J, Hubona G, Ray PA. [2016] Using PLS path modeling in new technology. 116(1): 2-20.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

A NOVEL APPROACH FOR PERFORMANCE AND ESTIMATION OF STUDENTS BY USING HYBRID MODEL

M. Saravanan[1*], V.L. Jyothi[2]

[1]*Sathyabama Institute of Science and Technology, Chennai, INDIA*
[2]*Jeppiaar Engineering College, Chennai, INDIA*

ABSTRACT

Now a day's huge volume of nodes using in educational system that contains some meaningful information's to predict students' performance. One of the data mining technique in this work is C4.5 algorithm. We have gathered for evacuation of undesirable data. In view of the grouping rules understudy dropout and disappointment is being anticipated. By utilizing every single accessible element, the examinations are led for enhancing the precision to anticipate which understudy has forecast of understudies. C4.5 is the prominent choice tree classifier in information mining. Exactness of this characterization calculation is contrasted all together with check best execution. After tree constructing the positioning of the understudy is figured on the premise of the understudy's inner evaluation. And after that the regular examples are created by utilizing FP development calculation.

KEY WORDS
Prediction, Machine Learning, C4.5, Educational, Performance

Received: 1 Nov 2018
Accepted: 9 Jan 2019
Published: 10 Jan 2019

INTRODUCTION

Instructive information mining is wide subject that gives machine learning, measurable data and also various types of information mining calculation to discover instructive datasets. Schools and universities have need to judge the scholastic productivity of understudies by evaluations or outside and inner imprints [1].Future subtle elements, for example, transporter alternative are anticipated about understudies utilizing various types of expectation models and likelihood of young people to increase merciless future For this reason various types of strategies, for example, grouping, affiliation information mining and differing characterizations are utilized. In proposed framework distinctive sorts.

Instructive information mining is as of late created slant and intriguing technique that gives different expectations in every single instructive level. Various techniques for information mining are exhibited following.

A. Estimation of Students scholastic execution
B. Foreseeing School dropouts
C. Understudies behavioral forecast

A. Estimation of Students scholastic execution
We display data examination of datasets to foresee the understudy's scholarly stamps and additionally understudy's positions were expert relies upon past record. To increment in the scholastic exhibitions of graduate understudies, we furthermore gave various sorts of information mining strategies, for example, grouping, affiliation govern mining and order and exception identification.

B. Anticipating School dropouts
A strategy furthermore proposed for checking rundown of understudies who dropouts the school. In this differing qualities, for example, participation, family foundation and sexual orientation decided for information mining [2]. Expectation of dropout is additionally finished with utilization of choice tree.

C. Understudy's behavioral forecast
With the help of proposed algorithm, calculation we show the change in understudy's execution and additionally association with employees and demeanor for foreseeing their conduct [3].

MATERIALS AND METHODS

Investigation of the condition of craftsmanship with identified with EDM and perception of the solid execution of this sort of date. Each understudy is characterized on the premise of kind of information and DM techniques and settled through sort of instructional occupation [4]. Creator investigated utilization of information digging in preparing for understudy's profile [5] and accumulations. Creator utilized apriori calculation of information digging for understudy's profile. K-implies bunching calculation utilized for understudies for exchange an arrangement of examination inside subset. Creator furthermore shows utilization of information digging technique for characterization and serves to understudies in determination of UG programs [6]. This paper moreover investigates examination on instructive structure and base of information mining process gathering the understudy's data's [7]. We preprocess the data we gathered for erasure of undesirable data. In view of the lead understudy dropout and frustration is being expected. In proposed work C4.5 count to suspect understudy frustration. Exactness of these portrayal figuring's is stood out all together from check best execution [8]. Understudy situating is done on the premise of understudy's interior evaluation. The positioning of understudy will be chosen by normal rate

***Corresponding Author**
Email:
mail2saravananme@gmail.com

computed and by arranging normal rate in slipping request [9]. And after that the successive examples are produced by utilizing FP development calculation the dataset contains 10 traits and 2 classes. The aggregate size of the dataset is 300[8]. It using C4.5 predict algorithm for throw predict result and identify how trained tutors use to improve the student performance with increase class rating [10]. All process output result is quality and accuracy. We propose c4.5 prediction analysis technique, this technique work as; include clustering result after analyzed student performance details. It given output is how much group of students only trained and after get good output result. That particular student database details refer and select tutors trainer for increase student performance [11]. Advantages of Proposed Systems are easily identify teacher decision in student performance, Accuracy and easy to maintain, providing a support and information for improving the college infrastructure and academic activities and students' performance based on their talent.

1: Preprocessing.
2: Calculate t= $\sum s(j/t)\log(s/t)$.
3: Ent of each trait $Li=Ent(Mi)=-\sum(s/Mi)/\log(s/Mi)$
4: Calculate each attribute $Ent(U) =-\sum(ni/n) * Ent(Mi)$
5: Build Tree.

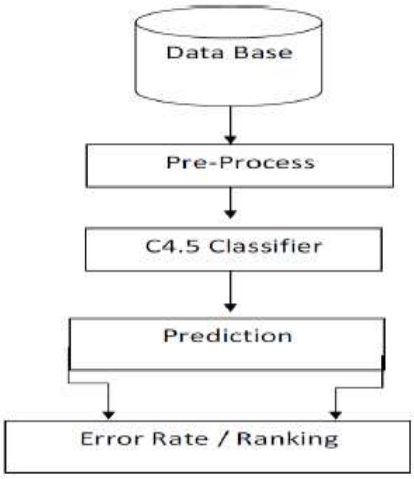

Fig. 1: Architecture of proposed method.

RESULTS

Two classes are utilized:
Great, AVERAGE
[Table 1] speaks to the examination of different classifiers with proposed calculation. Table unmistakably shows that the proposed calculation is superior to other classifier as far as precision. It effectively groups the occasion's precision

Table 1: Results for Comparisons

Algorithm	Training Set	Correctly Classified	Incorrectly Classified	Accuracy %
ADT	300	292	8	96
NB TREE	300	295	5	97.5
REP TREE	300	291	9	95.5
USER CLASSIFIER	300	211	89	55.5
PROPOSED ALGORITHM	300	297	3	98.5

Two classes are utilized:
Great, AVERAGE
Table 1 speaks to the examination of different classifiers with proposed calculation. Table unmistakably shows that the proposed calculation is superior to other classifier as far as precision. It effectively groups the occasion's precision. As showed up in [Table 1] the C4.5 classifier is used to process the results. Here the 200 get ready tests are used. While finding out, the total examples are separated into number of get ready and testing sets. What's more, after that the exactness is figured.

IIOAB JOURNAL
ISSN: 0976-3104

FP Growth happens

FP Growth is the essential count use to make Association rules. FP development is an approach in view of partition and vanquishes strategy. The principle reason behind this system is to deliver visit thing sets by utilizing the mix of information properties. It fundamentally deals with to create visit thing set without competitor set era.

Produced visit designs

The regular examples are produced on edge esteem 2. Just those examples are viewed as whose having bolster number 2 or >= 2. Furthermore, those examples are the regular examples.

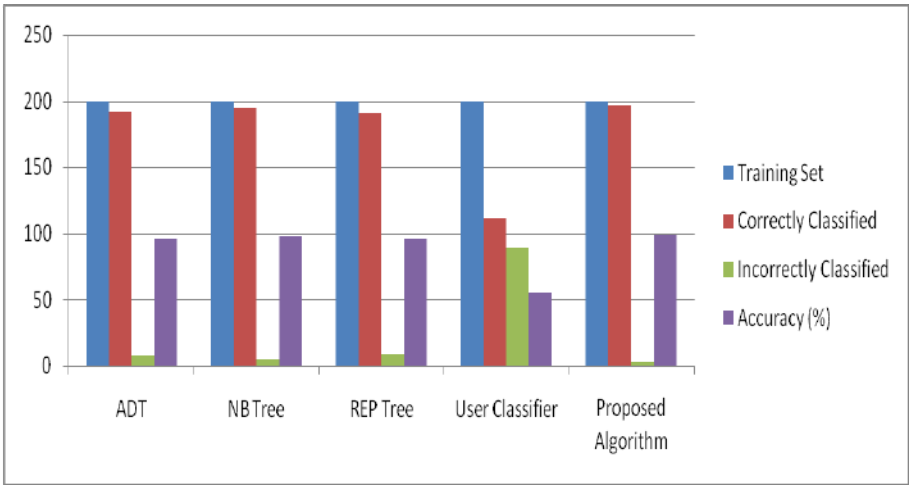

Fig. 2: Performance range.

CONCLUSION

Arrangement in information mining is wide region which draws in the specialists and correct and viably deal with the hunt of data. This paper shows grouping techniques to propose very much acted transporter for understudy. Undisciplined and fierce understudy influences their bearer. Order rules created by choice tree are renowned because of simple translation. Different sorts of classifiers are striven for figuring of exactness and execution and very much carried on classifier is chosen. In this way, probability of the understudy end up noticeably savage in future forecast is refined.. The execution of C4.5 classifier is measured as far as effectively arranged cases and erroneously grouped occasions. This expectation accommodating for establishment sorts out directing to fitting understudy on the premise of assessment of viciousness in starting stages. Different sorts of orders are used as prescient instrument inside information mining and thought about exhibitions. After that the positioning of the understudy is ascertained on the premise of understudy scholarly appraisal. And after that the successive examples are created by utilizing FP development calculation.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENTS

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Bin Mat U, Buniyamin N, Arsad PM. [2013] An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED):126–130.

[2] Ibrahim Z, Rusli D. [2007] Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression, in21st Annual SAS Malaysia Forum, 6 (20):12-16.

[3] Romero C, Ventura S. [2010] Educational data mining: A review of the state of the art, Trans. Sys. Man Cyber Part C 40 (6):601-618.

[4] Angeline DMD. [2013] Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1(1) :12–16.

[5] Saravanan M , Jyothi VL. [2016] The Role of Stress, Anxiety and Psychological Factors in Influencing the Performance of Students with an Analytical Approach of Classification

Method, International Scientific Researches Journal 72(9):142-46

[6] Jain LC, Raymond AT. [2007] Evolution of Teaching and learning paradigms in intelligent environment. Vol. 62. Springer.

[7] Bennedsen J, Caspersen ME. [2008] optimists have more fun but do they learn better? On the influence of emotional and social factors on learning introductory computer science.

[8] Fire M, Katz G, Elovici Y, Shapira B, Rokach L. [2012] Predicting Student Exam's Scores by Analyzing Social Network Data , chapter, Active Media Technology, Volume7669 of the series Lecture Notes in Computer Science:584-595.

[9] Nettleton DF. [2013] Data mining of social networks represented as graphs. Computer Science Review, 7:1-34.

[10] Romero, Cristobel, Sebastian Ventura [2007] Educational data mining: A survey from 1995 to 2005, 33(1):135-146.

[11] Ting SMR. [2001]Predicting Academic Success of First Year Engineering Students from Standardized Test Scores and Psychosocial Variables, 17(01): 1-8.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

PREDICTION OF PRETERM BIRTH USING DATA MINING-A SURVEY

Prema N. S.[1*], Pushpalatha M. P.[2]

[1 *]*Department of Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, INDIA*

[2]*Department of Computer Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, INDIA*

ABSTRACT

Background: *Preterm birth (PTB), the birth of infant before 37 weeks of gestation and it is the leading reason for perinatal mortality and morbidity. The reason for PTB is unclear, although many data mining techniques are used in identifying the risk factors. The reasons may be because of large and high dimensional data set. Many data mining (DM) methods are used to explore the risk factors of PTB and to predict preterm birth. In this paper, we describe the work carried out on application of data mining techniques towards the analysis and prediction of preterm birth.* ***Methods:*** *This survey paper is prepared by considering two criteria one the data mining techniques used for analysis and prediction and the other is the risk factors considered for the preterm birth prediction. Since in the works considered for survey the authors have not used a common or standard data set for the study/analysis, the comparisons of the works were difficult, here we have made an attempt to explore the data mining techniques for preterm birth prediction and the major risk factors of preterm birth.* ***Conclusion:*** *The most commonly used data mining technique is classification in that the usual techniques are Support Vector Machine (SVM) and Logistic regression. The most commonly considered risk factors are socio-demographic, behavioral (life style) and Pregnancy History.*

KEY WORDS
Data mining, Preterm birth, Logistic regression, Naive bayes, SVM, Neural networks, Decision tree

INTRODUCTION

Preterm birth is one of the public health risks in the society causing death and complications during pregnancy. A delivery that occurs twenty weeks after the start of labor and before 37 weeks of pregnancy is called preterm birth [1]. WHO reports that, about 15 million babies are born prematurely every year around the world and that is more than one in 10 of all babies born globally. In India, 3,341,000 preterm cases in every year and 361,600 children below five years die because of preterm complications [2]. The cause of preterm birth is complex, multi-factorial and not fully understood [3].

Types of preterm births

i. Spontaneous: This may occur after spontaneous on set labor or by prelabour premature rupture of membrane (PPROM). The cause of this type of preterm birth is very difficult to identify in up to half of all cases.
ii. Provider initiated: It is type of elective or induction labor before completion of 37 weeks of gestations for different reasons like fetal or maternal condition or other medical reasons.

Most occurring are spontaneous preterm births but some are due to early induction for medical or non-medical reasons. About 45-50% of preterm births are because of unknown reasons/causes, 30% are due to premature rupture of membrane(PROM) and other 15-20% is elective or medically indicated preterm deliveries.

Received: 30 Oct 2018
Accepted: 15 Dec 2018
Published: 10 Jan 2019

The main Risk factors for preterm birth are

i. Age of the women at pregnancy.
ii. History of preterm birth.
iii. Multiple pregnancies.
iv. Chronic diseases such as diabetes, hypertension, anaemia, asthma, thyroid disease.
v. Infection.
vi. Genetic influences.
vii. Nutritional factors: Under nutrition, obesity, nutritional deficiencies.
viii. Life style-women: smoking, consumption of alcohol, amusing drugs, Stress, excessive physical work [4].

Medical data mining

Data mining is the method of extracting useful knowledge from large repository of data. Medical data mining is an application of data mining, where data mining methods used for the analysis of medical data. The Medical data mining approaches are applied for the following tasks: treatment, prognosis, diagnosis, management and monitoring. The aim of medical data mining is to help and support physicians, care patients and improve public health.

The two main approaches of data mining are prediction and description. Prediction includes classification and regression,description includes clustering and association analysis.

In this paper, we describe work carried out on application of data mining techniques towards the analysis and prediction of preterm birth. We mainly considered the two things one is with respect to the data

***Corresponding Author**
Email:
prema.gowda@gmail.com

IIOAB JOURNAL
ISSN: 0976-3104

mining techniques used for the prediction and other is the risk factors /attributes considered for prediction.

RESEARCH PROGRESS

Many data mining models were developed for the prediction of PTB and the performances of the models are measured in terms of accuracy, AUC or ROC etc. The following are the models developed by various researchers.

A data mining model was developed on the raw data to identify the risk of PTB, monitoring and to control the PTB associated problems. The DM methods are applied on five different scenarios by combining different variables and also oversampling is done to balance the data by replicating the PTB cases so that the dataset reaches equal distribution of positive and negative instances. The risk factors considered for preterm birth prediction are pregnancy characteristics, Gestation and physical conditions of pregnant women [5].

Batoul ahadi et.al used statistical methods; support vector machine with different kernel functions and logistic regression models for preterm birth prediction. The data used is collected in 3 parts during nine months of pregnancy in every three months using questionnaires. The risk factors considered are demographic and pregnancy characteristics. They have applied Wrapper feature subset selection approach for the selection of best features [6].

The prediction model for PTB was developed by implementing risk of preterm delivery by Creasy [7], on Maternal-fetal Units Network (MFMU) data set. The data set used includes all most all risk factors of preterm birth in different time points Because of large size of the data set they have done preprocessing of data to remove noise, handled missing values by combining attributes into different groups. The authors conclude that identifying the risk factors of PTB is not elusive it requires an efficient model for correct prediction. They have mainly focused on first time mothers (nulliparous women) because of unavailability of pregnancy history [8]. In their further work they have demonstrated that model selection and nonlinear kernel methods for the prediction of PTB are better approaches for promising results [9].

Tuyen Tran et.al have presented methods for preterm birth prediction which includes quantifying, discovering risk factors, derivation of interpretable prediction rules and usage of stabilized sparse logistic regression for deriving linear prediction models. The authors also used Randomized Gradient Boosting a hybrid model to estimate the upper-bound accuracy for the data [10].

Cluster analysis techniques were used to get mutual exclusive clusters, as there were very low similarities in the set of attributes; they went for classification techniques, the accuracy of the classifiers are improved by applying cross validation. The classification techniques used are SVM and Naïve Bayes with the highest accuracy of 90% and 88% respectively [11].

Hsiang-Yang Chen et.al has explored the major risk factors of preterm birth using neural networks and decision tree. A neural network is used to find the top 15 risk factors then they have used decision tree C5.0 for classification. They have achieved an average accuracy of 80% and their results shows that hemorrhage in pregnancy and multiple birth are the major risk factors for preterm birth. Further they have also considered paternal risk factors like drinking alcohol, smoking and occupation in their study [12].

The prediction of PTB was done by using associative classifier by Yavar Naddaf et.al, they have also applied number of classification methods on a data set by considering both maternal and fetal records in predicting preterm birth. The Performance of the methods is very poor and there is no much improvement in the performance after applying feature selection techniques [13].

Christina Catley et.al developed an artificial neural network(ANN) model for the prediction of preterm birth on Perinatal Partnership Program of Eastern and South eastern Ontario (PPPESO) database based on physician input, eight obstetrical variables namely Age, Current pregnancy number of babies, Parity, Baby's gender, intention of breastfeed, Smoking, Previous term and preterm babies. The authors also shown the effect of change in the prior distribution of data on the performance of back propagation feed-forward ANN and they have assessed the effectiveness of weight elimination cost function in improving the accuracy [14]. In their further work they have applied ANN for prediction of high risk preterm birth. The back propagation feed forwarded ANN was developed by considering variables describing pregnancy history for high risk preterm birth prediction, the performance is measured in terms of sensitivity (36%&37%) and specificity (88% &92%) [15].

M. M. Van Dyne et.al used Learning from Examples using Rough sets (LERS) where rules are generated directly from the data for preterm birth prediction. The LERS model was able to predict 78% correctly of full term cases and 90% of preterm cases when full term and preterm rules were run separately and 73% of accuracy for combined cases [16].

Linda Goodwin and Sean Meher used the following techniques for preterm birth prediction namely neural networks, logistic regression, CART and software (PVRuleMiner and FactMiner) there was very small differences in the performances of the techniques. The data set used for the study is from Duke University

IIOAB JOURNAL
ISSN: 0976-3104

medical center data repository, they have considered about 32 demographic parameters for preterm birth prediction [17].

The usage of data mining techniques for preterm birth prediction was shown by considering about seven demographic parameters. The obtained results were interesting but the concern is whether the particular demographic data selected would represent the general population [18].

A secondary analysis was done by Courtney et al. describing that the preterm prediction model generated in [18] using demographic parameters is generalizing to a larger population with a modest result. In the study the data set used is birth certificate data, the demographic parameters considered for prediction are only five namely Age, marital status, ethnicity/race, education and country. The acceptable average AUC result for classifiers is 0.58 for different mining techniques [19].

In our work we have also used the statistical models for the prediction of PTB; the main risk factors considered are Age, Number of Times Pregnant, Diabetes, Obesity and Hypertension. The data set considered the details of pregnant women having either diabetes mellitus or gestational diabetes mellitus. The data set was balanced by using data oversampling techniques Synthetic Minority Oversampling Techniques (SMOTE). Comparing with work carried out by other researcher's highest accuracy achieved but the risk factors considered are limited and data sample size is also small [20].

Table 1: Data mining methods and accuracy for preterm birth prediction

Sr. No.	Data Mining Methods	Performance measures	Reference	Year
1	Decision Tree Generalized Linear Model Support Vector Machine Naïve Bayes	93% 86% 93% 74% (Accuracy)	[5]	2015
2	Support Vector Machine Logistic Regression	56% 67% (Accuracy)	[6]	2016
3	Support Vector Machine (Linear, Polynomial, RBF) Logistic Regression(Lasso, Elastic Net)	60% (Accuracy average)	[8]	2016
4	Logistic Regression Randomized Gradient Boosting(Ensemble Method)	62%/81.5% (Sensitivity/Specificity)	[10]	2016
5	Hierarchical Clustering Naïve Bayes Support Vector Machine	- 88% 90% (accuracy)	[11]	2010
6	Neural Networks & Decision Tree C5.0	80% (accuracy)	[12]	2010
7	Logistic Regression Naïve Bayes SVM Neural Networks Decision Tree C4.5 Associative Classifier	0.57 (Average AUC)	[13]	2006
8	ANN	36%/90% (Sensitivity/Specificity	[15]	2006
9	Neural Networks CART	0.64 0.65 (ROC)	[17]	2000
10	Logistic Regression Neural Networks SVM Bayesian Classifier CART	0.605 0.57 0.57 0.59 0.56 (AUC)	[19]	2008
11	Support Vector Machine Logistic Regression	87% 86% (Accuracy)	[20]	2018

Table 2: The most common Risk factors considered for preterm birth prediction

Factors Contributing to Preterm Birth	Variables	References
Demographic and Socioeconomic	Maternal Age Race/Ethnicity Educational Status Maternal Status Socioeconomic Status	[13] [8] [9] [11] [5] [12] [17] [18]
Behavioral Characteristics/ Life style	Alcohol Tobacco Recreational Drugs Psychological And Social Stress	[12] [11] [5] [8] [9] [10]
Maternal Health/ Chronic conditions	Body Mass Index (BMI) Diabetes Hypertension Anemia Asthma Thyroid Disease	 [5] [13] [20]
Current Fetal Conditions/Pregnancy Characteristics	Multiple Fetuses Infertility Treatments Infant weight Drugs used during pregnancy	 [5] [10] [6]
Pregnancy History/ Genetic Characteristics	Previous Preterm Births Diabetes Hypertension Obesity	[8] [9] [10] [13]
Biological Characteristics	Infections	[8] [9]
Others	Ultrasonography Insurance Details Cervical Measurements	 [5] [8] [9]

CONCLUSION

In this paper we have made an attempt study the work related to preterm term birth prediction using data mining techniques. The work considered for survey are studied based on the risk factors considered for prediction and the data mining techniques used, the comparative study was difficult as none of the authors have considered either the dataset or the risk factors in common.

The following conclusions can be drawn after the study of various works done in the application of data mining in preterm birth prediction.

1. Data set: All the works carried out in literature have not considered any data set in common for the comparison of performance of their proposed data mining techniques. All most all have used different data sets with varying size and dimensions.
2. Data sharing and privacy: As the work related to personal medical details of individual which will be circulated as commercial product which would help in research work in healthcare industry but at the same time it might threat privacy protection.
3. DM methods: The most commonly used data mining method is classification for predicting preterm birth and also for exploring risk factors of preterm birth. The very regularly used classification techniques are Support vector machine with different kernel functions and Logistic regression. Very few authors have used clustering and Association analysis techniques for preterm birth prediction but their performance is very poor compared to that of classification methods.
4. The main risk factor identified for the prediction is previous preterm birth but it is difficult in first time mothers because of unavailability of pregnancy history.
5. Performance measures: The classification techniques performances are measured in terms of accuracy, sensitivity/specificity ROC and AUC the details are shown in [Table 1].

6. Major risk factors: The different risk factors considered for PTB are shown in [Table 2]. The most commonly considered risk factors are socio-demographic, behavioral (life style) and Pregnancy History.

FUTURE WORK

In literature few works can be found where data mining techniques are applied for the prediction of preterm birth using Electro-hysterography(EHG) signal, EHG is used to measure electrical activity in the uterus. In this survey we have not included preterm birth prediction using EHG, which can be included in future.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENTS

The authors express gratitude towards the assistance provided by Accendere Knowledge Management Services Pvt. Ltd. in preparing the manuscripts. We also thank our mentors and faculty members who guided us throughout the research and helped us in achieving desired results.

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Cunningham F, Leveno K, Bloom S, Hauth J, Rouse D. Spong CY.[2010] Williams ObstetricsUSA : The McGraw-Hill Companies, Inc. Medical Publishing Division.

[2] WHO. [Online] //www.who.int/topics/preterm_birth.

[3] National Health portal.[Online] www.nhp.gov.in/disease/reproductive-system/female-gynaecological-diseases-/preterm-birth

[4] Varney H, Kriebs JM, Gegor CL.[2004] Varney's midwifery: Jones & Bartlett Learning. Canada.

[5] Pereira S, Portela F, Santos MF, Machado J, Abelha A.[2015] Predicting Preterm Birth in Maternity Care by means of Data Mining. Progress in Artificial Intelligence 116-121.

[6] Batoul A, Hamid A, Soheila K, et al.[2016] Using support vector machines in predicting and classifying factors affecting preterm delivery. Journal of Paramedical Sciences. 7(3): ISSN 2008-4978.

[7] RK Creasy, BA Gummer, GC. Liggins.[1980] System for predicting spontaneous preterm. Obstetrics and Gynecology, 55(6):692-695.

[8] Ilia Vovsha, AnsafSalleb-Aouissi, Anita Raja, et al.[2016] Predicting Preterm Birth Is Not Elusive: Machine Learning Paves the Way to Individual Wellness. Proceedings of the 1st Machine Learning for Healthcare Conference, PMLR 56:55-72.

[9] lia Vovsha, Ashwath Rajan, Ansaf Salleb-Aouissi, et al. [2014] Predicting Preterm Birth Is Not Elusive: Machine Learning Paves the Way to Individual Wellness.Stanford University : Big Data Becomes Personal: Knowledge into Meaning :AAAI Spring Symposium Series.

[10] Tran T, Luo W, Phun D, Morris J, Rickard K. [2016] Preterm Birth Prediction: Deriving Stable and Interpretable Rules from High Dimensional Data. Proceedings of Machine Learning for Healthcare 2016 JMLR W&C Track (56).

[11] Adriana-Georgiana MALEA, Ştefan HOLBAN, Nicolae MELIȚĂ.[2010] Analysis and Determination of Risk Factors using R. 10th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS, May 27-29

[12] Hsiang-Yang Chen, Chao-Hua Chuang ,Yao-Jung Yang , Tung-Pi Wuc[2011] Exploring the risk factors of preterm birth using data mining. Expert Systems with Applications.38 (5): 5384-5387.

[13] Yavar Naddaf, MojdehJalali Heravi and AmitSatsangi. [2008]. Predicting Preterm Birth Based on Maternal and Fetal Data. semanticscholar.

[14] C Catley, M Frize, RC Walker, DC Petriu. [2005] Predicting preterm birth using artificial neural networks. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05).1064-7125.

[15] Catley C, Frize M, Walker RC, petriu DC.[2006] Predicting High-Risk Preterm Birth Using Artificial Neural Networks. IEEE Transactions on Information Technology in Biomedicine, July 2006 , 10:540-549.

[16] Dyne MM Van, et al., et al.[1994]. Using Machine learning and expert systems to predict preterm delivery in pregnant women. USA : Tenth Conference on Artificial Intelligence for Applications.

[17] Goodwin L, Meher S.[2000] Data mining for preterm birth prediction. Proceedings of the 2000 ACM symposium on Applied computing, 46-51.

[18] Linda K Goodwin, Mary Ann Iannacchione, et al.[2001] Data mining methods find demographic predictors of preterm birth. Nursing research, 340-345.

[19] Courtney KL, Stewart S, Popescu M, Goodwin LK. [2008] Predictors of preterm birth in birth certificate data .Studies in health technology and informatics, February.136:555-560

[20] Prema NS, Pushpalatha MP. [2018] Machine learning approach for Preterm Birth Prediction Based on Maternal Chronic Conditions. International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT-2018).

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

A DETAILED STUDY OF VARIOUS CHALLENGES IN CLOUD COMPUTING

Arushi Garg[1*], Shruti Suman[2], Mansi Goyal[3], Veena Tayal[4], Prateek Jain[5]

[1,2,3,4] *Department of CSE, Manav Rachna International Institute of Research & Studies, INDIA*

[5] *Accendere Knowledge Management Services Pvt. Lt. INDIA*

ABSTRACT

Cloud computing is internet based ("cloud") development and use of computer technology ("computing"). Because of accessibility of diverse services and extensibility for vast areas of computing processes, individual users and organizations convey their data and services to the cloud storage server. The point of demarcation between the responsibilities of the provider and those of the user is denoted by the cloud symbol. The network infrastructure as well as the servers are covered in cloud computing since they lie within its boundary. A set of IT services provided to a customer over a network on lease with the ability to scale up or down their service requirements is called cloud computing. Cloud computing services are usually delivered by a third-party provider who owns the infrastructure. The security issues and challenges associated with cloud computing are making organizations hesitate in accepting it despite potential gains achieved. Nevertheless, of its advantages, the transformation of local computing to remote computing has brought up many security issues and challenges for both the users and providers. One of the major issues that hamper the growth of cloud is security. This paper focuses on existing issues in cloud computing such as security, privacy, reliability and so on.

KEY WORDS
Cloud Computing, Challenges, Issues, Models, Security

Received: 15 Nov 2018
Accepted: 24 Dec 2018
Published: 12 Jan 2019

INTRODUCTION

Cloud computing is a representation for empowering appropriate, on demand network approach to a shared pool of arrangement of computing resources that can be swiftly planning and unconfined with minimal administration effort or service provider communication [1].

Cloud Computing is a different method of commercial computing. It will be extensively used in the nearby future. The fundamental notion of cloud computing is dropping dispensation encumbrance on the customer's terminal. These are obtainable via a fast internet connection. This is a whole new concept that many organizations are adopting as to give their customers more memory space to keep their files. Although this is a sedition technology used worldwide numbered third most used latest technology, this also signifies the growing trend in the market nowadays in IT industry.

In specific, five vital features of cloud computing are clearly articulated in [Fig.1].

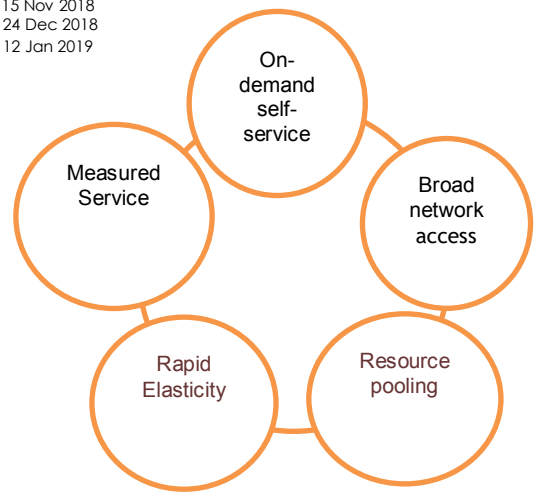

Fig.1: Vital Features of cloud computing.

- On-demand self-service

A user with a prompt necessity at a precise timeslot can benefit computing resources in an involuntary (i.e. convenient, self-serve) manner without resorting to human communications with providers of these properties. [1]

- Wide-ranging network access

Cloud computing properties are brought over the network and used by numerous applications with mixed platforms located at customer's website.

***Corresponding Author**
Email:
arushigarg1997@gmail.com

IIOAB JOURNAL
ISSN: 0976-3104

- Resource Mutualisation

Cloud service provider's resources are 'mutually together' in a determination to aid various customers using virtualization model or perhaps software multi-tenancy.

- Quick elasticity

Resources are instantaneous not persistent, there will be no agreement and in conduct so as to they can be used to scale up whenever they want and scale down to release them.

- Measured Service

As computing resources are mutual and communal by various customers, cloud substructure uses suitable to verify the procedure of the numerous machineries to amount the procedure. [1]

CLOUD COMPUTING AND BUILDING BLOCKS

Service model

There are 4 types of services:

- *Software as a Service (SaaS)*

SAAS is the huge area that holds all the cloud services it provides the customer. It is referred to as "On-demand Software". SaaS is utilized by users through web browser using lightweight computer that has been modified into server –based computers. It achieves standardizationwith respect tovelocity, cost effectiveness, securing the data, availability of data, maintaining it too.Examples are Google Mail, Google Docs, etc. [2]

- *Platform as a Service (PaaS)*

PaaS comes under category of cloud computing services that provide platforms to users for them to develop, run and monitor applications without the difficulty of building and maintaining infrastructure which includes developing and launching an app. So, if we differentiate between these two services it will result in giving solution as- SaaS will only launch fully develop Applications whereas PaaS will lunch both fully developed and in-progress cloud applications. It can be delivered in 3 ways:

  – private service,
  – public service, and
  – private and public software.

Example is Google AppEngine. [2]

- *Infrastructure as a Service (IaaS)*

IaaS is a form of cloud computing service that provide virtual resources to users over the internet. A hypervisor runs the virtual machines as guest machine. It provides high-level APIs for dereferencing various network infrastructure. IaaS is the capability provided to user to provision various fundamental resources so 14 user can run these software.
Example is Amazon's EC2. [2]

Deployment model

Recently, new deployment models have been introduced in Cloud Computing.

- *Private cloud*

The cloud organization is exclusively functional with a single association and is managed by other bodies irrespective of the foundation it is set upon.
The reason behind developing a private cloud are quite many. Following are the reasons:

  – Increasing the utility of in-house resources
  – Security purposes
  – Cost for data transfer from IT infrastructure to public cloud is considerable
  – Academics play a major role
  – Organization require full access of all activities going on in cloud system. [3]
  –

- *Community cloud*

Community cloud refers to sharing of same cloud infrastructure and also its policies behind joining the cloud. It is a viable option as:

  – Economic scalability
  – Symmetry in form of creating policies.

This cloud can be presented through third party bodies or IT in the following communities. [3]

- *Public Cloud*

Public Cloudis the most used type of deployment model of cloud computing. This type is utilized by many customers. The cloud service provider holds full ownership to public cloud and has its own policies, profit, charging etc. Examples are Amazon EC2, S3, Google AppEngine, and Force.com [3].

IIOAB JOURNAL
ISSN: 0976-3104

- *Hybrid cloud*

This cloud model is a combination of two or more clouds by allowing sharing of data and application with other clouds [3].

CHALLENGES IN CLOUD COMPUTING

In previous decades to pass, cloud computing has proved to be a promising business concept the IT industry all over the world currently. Now, IT companies are realizing that by entering the world of cloud computing, they can make most of the profit in their business at a negligible price. But as we know that every advantage has a con tail, which is why in spite of having millions of advantages it lacks the following features that every IT company dreams to have in their cloud computing package in [Fig. 2] [4].
Challenges:

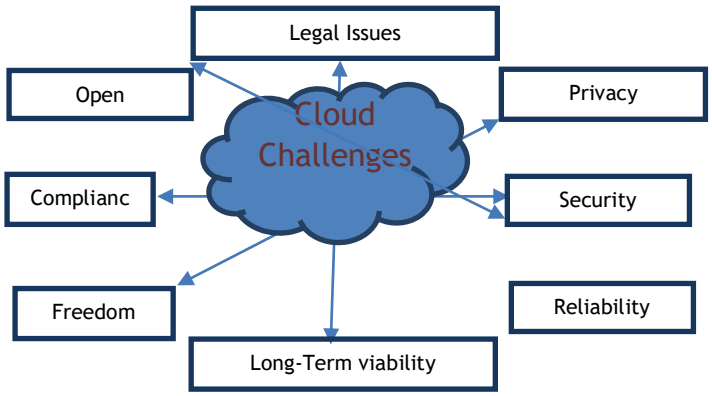

Fig. 2: Challenges in Cloud Computing.

- **Security**

Is your data more secure on local hard drive or on high security servers in the cloud that IT companies sell? Some might argue that the data is secure more if it is managed internally, in other hands some says that cloud provides high tech security to users. However, the data that is in cloud is stored in multiple computers. So how is he data safe in cloud? If the hacker access any of these computers, he/she might be able to use our information in any form he/she wants. The IT companies that provide cloud platform must work on providing high security to their customers in order for them to rapidly increase their reputation in the business world [4,5].

- **Privacy**

Customer's personal data might be distributed among different computers rather than stay in one physical location. This might lead to violation of privacy of customer. It may even be possible that the customer may leak information when using the services of cloud [4,5].

- **Reliability**

Cloud servers and resident servers have the exact problems. It means, that the cloud server experiences downtime, interruptions and slowdowns.The only difference is that customer has much larger dependencies on cloud service providers in the cloud computing model [6].

- **Performance and bandwidth cost**

IT companies need to spend more on bandwidth if not on hardware. This will result in low cost for smaller applications but may be quite high foe data-intensive application. We require sufficient bandwidth to deliver intensive and complex data over the network. Due to this, many companies are in line waiting for a reduced cost before switching to cloud [6].

- **Open Standard**

Open standards are very important for viable increase in the growth of cloud in IT companies. Most cloud providers reveal APIs which are well documented and unique too which makes it non-interoperable. Some uses others APIs which are under open standards [6].

- **Compliance**

Using of data requires audit and reporting trails, CSP must allow user to fulfill agreements and regulations. If we manage protection and compliance, it deliversvisionon how an opinionof IT Company can transfer a tight Management and fulfillment of compliance regulations. In conclusion, the customers' requirements are maintained by cloud providers [7].

- **Freedom**

Cloud providers doesn't enable customers to access the data storage. It means that data controls are for cloud computing.Users tend to deny when not given freedom to retain their duplicate copies of data to retain their choice to freedom and secure them too [7].

- **Long-term Feasibility**

One must ensure that all of the data one must put up to not perish. If, so ask your cloud provider as to how can you search the retrieval data back and will be replaceable setup [6,7].

DATA STORAGE AND SECURITY IN CLOUD COMPUTING

We can define cloud security as protection of all the data related to cloud provider as to no unauthorized user can hack it. It is a broader field of security of computer, network and information [8,9].

The major services provided such as a shared resource, identity management, privacy and access control are particular concern. Since more companies are now switching to cloud-based services and the related providers for operations of data, security in vulnerable areas to become priority for cloud computing providers. A study by current review displays that securing of data and privacy of risks are becoming major concern for people who are shifting to cloud computing. Hence, we can say that many complications still exist in cloud computing. Contrarily, a distinct person will have control on security of data and processed on their computers. Additionally, the services and maintenance of data are provided via a vendor where no client/customer are unaware of how will all the processes are working. So, rationally talking, the user will have no control over it. If we talk of data security in cloud, the vendors provide Service Level Agreements to ensure assurance.

CLOUD SECURITY ISSUES

- **Data Security**

We should secure the data as it is main concern of all the companies as it may hamper their reputation if security is breached. Almost 70% survey tells that security are breached in these IT industries quite easily. [7][8]

- **Refusal of regulating security orders**

Approximately 40% of the IT industries are concerned as to how to make sure agreement of the company is achieved so that reputation of companies is achieved. If there is a security breach then it will hamper the compliance and result to fines and loss of business contracts. [8]

- **IT services not to control**

The survey indicates that their fear for loss of control in IT industry can manifest in many ways. Here, data moves to cloud services. [8]

- **IT expertise**

Almost 40% of industries are not controlling cloud platforms. The reason may be knowledge and IT expertise. [10]

- **Compromised accounts or insider threats**

30% of survey indicates which accounts are used by SaaS provider and can be compromised in quite methods which is not important here. [10]

- **Continuity in business**

If a company loses all the accesses of its IT industry, it may seem as it has gone out of business. It can be a rare scenario but it is 30% of the chances. [10] [11]

SECURITY ALGORITHMS

Every user needs a secure communication channel and large data storage over the network. For this purpose the encryption algorithm plays a vital role in the network. It is the fundamental tool for protecting the data. Encryption/Decryption algorithms convert the message into encrypted form using "the key" and only the authorized user or receiver has that key therefore, can decrypt the transmitted message using provided key. Thus it ensures confidentiality, integrity, availability and authentication of the data [Table-1]. There are two types of security algorithm; symmetric key encryption and asymmetric key encryption. [12][13][14]

Table 1: Key used for Encryption/Decryption

Key of comparison	DES	AES	RSA	MD5
Key used for encryption/Decryption	Same key is used	Same key is used	Different keys	Hash function is used for one way cryptography
Key Size	56	128, 192, 256	>1024	128
Performance	Slow	Fast	Fast	Faster

- ***RSA (Rivest-Shamir-Adelman)*** – Also known as public key algorithm. It is the mostly uses the algorithm- Asymmetric key algorithm, which consists of 2 keys that are:

– Public

– Private

We will use public key is used to encrypt data and is known to everyone and private key is not known to anybody except receiver. The server performs asymmetric key by encrypting a distinctive memo for example, digital signature to enforce authenticity of the sender [15].

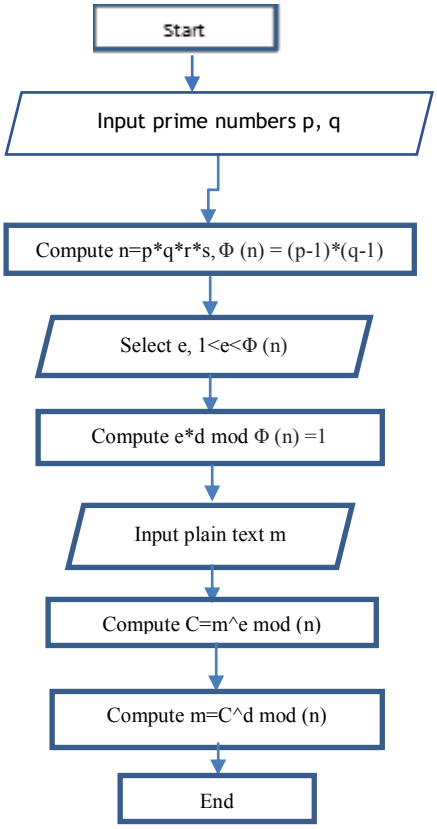

Fig. 3: RSA Algorithm [15].

- ***DES (Data Encryption Standards)***- DES is a symmetric key block cipher. It was published and accepted by the National Institute of Standards and Technology (NIST).The encryption process is made of two permutations and sixteen Fiestal rounds. DES uses 16 rounds .Each round of DES is a Fiestal cipher. It works by using the same key for encryption and decryption of message, so both the sender and receiver must know and use the same private key .DES uses 64-bit key ,and the length of the key determines the number of possible keys and hence feasibility of attack on the system.

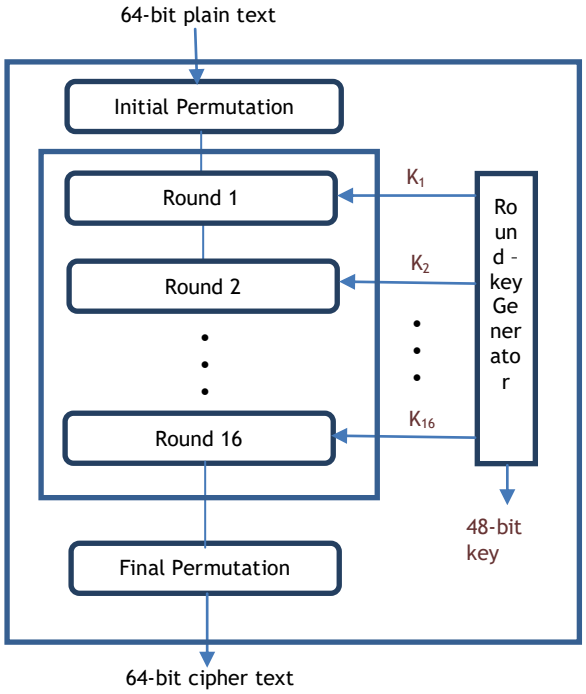

Fig. 4: Data encryption standard.

.....

- **AES** ***(Advanced Encryption Standard)*** - AES is a symmetric key block cipher which was published by the National Institute of Standards and Technology (NIST).It is also known as non-Fiestal cipher that can encrypt and decrypt a data block of 128 bits. It uses 10, 12, or 14 rounds. The key length, which can be 128, 192, or 256 bits, depends on the number of rounds. It generates the encrypted hash code in highly secure manner which also ensures confidentiality of the message. It works on three broad categories mainly:
  - Security
  - Cost
  - Implementation

  The AES replaced the DES with new and updated features. The techniques involved in this algorithm are so easy and flexible such that they can be easily implemented using cheap processors and also consumes less amount of memory. The symmetric algorithm requires only one encryption and decryption key. It ensures data security for 20-30 years. Allowing worldwide access with no royalties, it has overall easy implementation. [16]

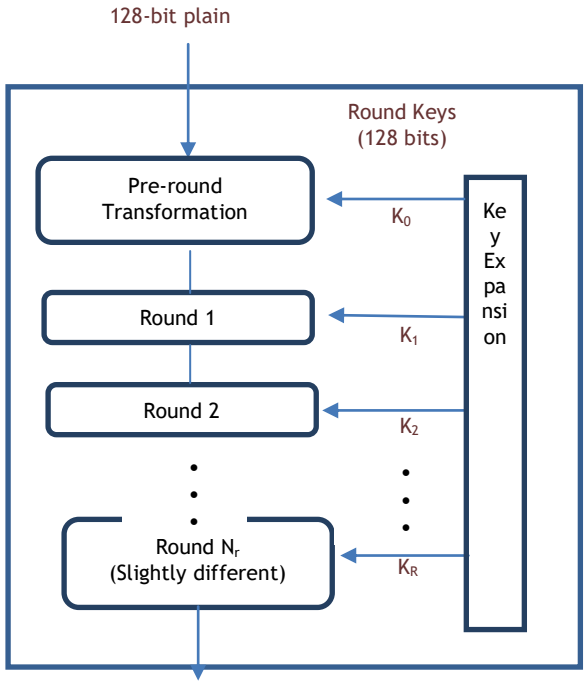

Fig. 5: Advanced Encryption Standards.

.....

- *MD5 (Message Digest Algorithm 5)*- It is the most commonly used cryptographic hash function algorithm with 128 –bit hash value and it can generate variable length message into fixed length output of 128 bits. Even though MD5 is prone to attacks, but still it can be used as a datum to verify and ensure the integrity and authentication of the message transmitted. In MD5 algorithm firstly, the input text provided is broken into chunks of 512 bits blocks and then the text is padded so that its total length is divisible by 512.

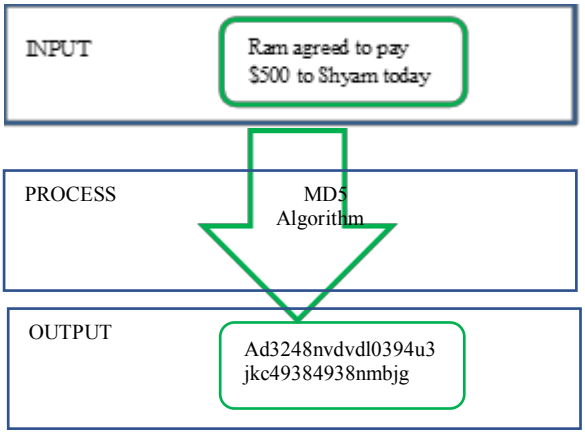

Fig.6: Message Digest 5(MD5).

……

RECENT TRENDS IN CLOUD COMPUTING

1) **Surge in Demand For Cloud Experts**
   - As many people are adopting to cloud, IT infrastructure requires more experts to become a major competition to its rival business.
   - So, IT manager searches for candidates that have gained experience in cloud platform as improving these skills is becoming a top priority.

2) **Hybridization of technology**
   - Many smaller companies or start-ups have moved to cloud platform permanently but it's a different scenario for larger companies to shift their entire IT operations to cloud based platform, security being an essential issue. [16]
   - So what these large organization do, they partially shift to cloud platform and rest is handled as it was which makes it complex.

3) **Reduced Software Restrictions**
   - This software puts an end to consumer's freedom. It infrastructure often restrict consumer's right to download and application permissions.
   - But the increase in cloud adoption, services allow IT infrastructure to control applications without neglecting user's choices.

4) **Rise of SaaS**
   - To gain successful operations, one must invest heavily in hardware, people and software too.
   - But again cloud adoption has proved to be beneficial as it cuts the operational costs without reducing the profit a business a can earn. [17]

5) **Increased focus on Long Term Relationships**
   Cloud platform has proved that it can run alongside the IT infrastructure and business for the long-term relations in order to gain customer's success.

6) **Rebranding of Service Providers**
   Before cloud was introduced, service providers focused on on-premises software's technical supports.

TECHNIQUES TO SECURE DATA IN CLOUD

- **Authentication and Identity**

Authentication of users and communicating systems is achieved by various means (Cryptography being the most common). Authentication of users happen via various methods [18][19]:

- Passwords, and
- Security token or fingerprint.

Identification of users can be achieved via:

- Fingerprint
- Passwords etc.
-

- **Data Encryption**

We need to use data encryption techniques if we are to store confidential information. Nowadays hackers can bypass passwords and firewalls in a few seconds to access the critical data. An encrypted data cannot be read until and unless it is decrypted with an encryption key.It is a technique of converting plain text into cipher text. Only the receiver will have an encryption key as to read the data provided to him. [18]

- **Information integrity and Privacy**

Cloud computing deliver information and resources to authorized users only. We can achieve integrity by providing communal faith between receiver and sender. We can also achieve authentication, authorization and accountability to ensure that the data is in hands of authorized user. We can secure our data by security Algorithms discussed above – RSA, DES etc. [19]

- **Availability of Information (SLA)**

Availability of resources, in terms of information security, means that the resources to an authorized user must be available whenever he/she needs. So, if the information is unavailable it creates a havoc. We use SLAs to provide information about whether the users have available resources they need or not. We can have a backup plan of resources and crucial information for a way to provide available resources. This will allow the user to have the information even if they are un-available. [19][20]

- **Secure Information Management**

It is the practice of gathering, monitoring and analyzing security related data from logs. A Security Information management System powers that practice. It is sometimes called security event management. It includes log data created from various sources like IDS, IPS routers, servers etc. SIMS will do following activities [20]

- Monitor events
- Display real time view
- Translate event data into common format, XML.
- Aggregate data
- Correlate data

- **Malware-injection attack solution**

Here, number of client virtual machines are generated and stored in a central storage location. Here we use File Allocation Table which consists of various virtual operating systems. This table consists of applications that are run by client. All the cases are monitored and run by hypervisor. We can use IDT for checking integrity. [20]

CONCLUSION

Security issues in the field of cloud computing are active area of research and experimentation. Various issues have been identified one of which is security of user data and applications. Security is always a major concern in open system architectures and needs to be managed effectively. Several cloud computing services are available to achieve security with varying techniques and methods. There are many new technologies emerging at a rapid rate, each with technological advancements and the potential to make human's life easier. However, one must be careful to understand the security risks and challenges posed due their utilization and implementation. In this study, we have discussed key security considerations and challenges which are currently faced in the cloud computing and also specified some techniques to secure data in the cloud. Owing to the issues which are prevalent, a proactive approach to the adoption of the Cloud computing is advised.

Adopting few security measures from user end, can go a long way in maintaining secure files on cloud. Cloud computing has the potential to become a frontrunner in promoting a secure, virtual and economically viable IT solution in the future. Cloud Computing security challenges are a part of ongoing research. Various open issues are identified as a future scope.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Dr. Prateek Jain, Accendere Knowledge Management Services., Ms. Veena Tayal, FET, MRIIRS for their valuable comments that led to substantial improvements on an earlier version of this manuscript.

FINANCIAL DISCLOSURE

None

IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] Stergiou C, Psannis KE, Kim B, Gupta, B. [2018] Secure integration of IoT and Cloud Computing. Future Generation Computer Systems, 78(2): 964-975.

[2] Kumar S, Goudar RH. [2012] Cloud Computing – Research Issues, Challenges, Architecture, Platforms and Applications: A Survey. International Journal of Future Computer and Communication, 1(4): 356-360.

[3] harma M, Husain S, Zain H. [2017] Cloud Computing Architecture& Services. IOSR Journal of Computer Engineering, 19(2):13-18.

[4] Singh A, Chatterjee K.[2017] Cloud security issues and challenges: A survey. Journal of Network and Computer Applications, 79:88-115.

[5] Deepthi S, Tulasi V. [2013] A Study of Security Issues and Cloud Models in Cloud Computing. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(12): 3265-3270.

[6] Zhou MQ, Zhang R, Xie W, Qian WN, Zhou A. [2010] Security and Privacy in Cloud Computing: A Survey, 2010 Sixth International Conference on Semantics, Knowledge and Grids(SKG), 105-112. Doi: 10.1109/SKG.2010.19

[7] Kumar S, Goudar RH. [2012] Cloud Computing – Research Issues, Challenges, Architecture, Platforms and Applications: A Survey. International Journal of Future Computer and Communication, 356-360.

[8] Dillon T, Wu C, Chang E. [2010] Cloud Computing: Issues and Challenges. 2010 24th IEEE International Conference on Advanced Information Networking and Applications. Doi: 10.1109/AINA.2010.187

[9] Vurukonda N, Rao BT. [2016] A Study on Data Storage Security Issues in Cloud Computing. Procedia Computer Science, 92:128-135.

[10] ICCSEA 2012, Wyld, D. C, Zizka, J, Nagamalai, D. (Eds.). [2012]. Advances in computer science, engineering & a4pplications: Proceedings of the second International Conference on Computer Science, Engineering & Applications (ICCSEA 2012), May 25-27, 2012, New Delhi, India. (Advances in computer science, engineering & applications.) Berlin: Springer.

[11] Stallings W. [2005] Cryptography and Network Security, Prentice Hall, 4th Ed, 2005.

[12] Khan SS, Tuteja, RR. [2015]. Security in Cloud Computing using Cryptographic Algorithms. International Journal of Innovative Research in Computer and Communication Engineering, 3(1): 148-154.

[13] Vijayapriya M. [2013] security algorithm in cloud computing: overview. International Journal of Computer Science & Engineering Technology (IJCSET), 4 (9): 1209-1211.

[14] Yang JF, Chen ZB. [2010] Cloud Computing Research and Security Issues. 2010 IEEE International Conference on Computational Intelligence and Software Engineering (CiSE), 10.1109/CISE.2010.5677076

[15] Pellegrini A, Bertacco V, Austin T. [2010] Fault-based attack of RSA authentication. 2010 Design, Automation & Test in Europe Conference & Exhibition. Doi: 10.1109/DATE.2010.5456933

[16] Sumitra. [2013] Comparative Analysis of AES and DES security Algorithms , International Journal of Scientific and Research Publications, 3(1): 1-6.

[17] Suresh KS, Prasad KV. [2012] Security Issues and Security Algorithms in Cloud Computing, International Journal of Advanced Research in Computer Science and Software Engineering, 2(10): 110-114.

[18] Jakimoski K. [2016] Security Techniques for Data Protection in Cloud Computing. International Journal of Grid and Distributed Computing, 9(1):49-56.

[19] Angadi AB, Angadi AB, Gull KC. [2013] Security Issues with Possible Solutions in Cloud Computing-A Survey, IJARCET, 2(2): 302-311.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

COMPARATIVE STUDY ON WEB SERVICE DATA INTERCHANGE FORMATS

Vishawjyoti*, Rashmi Agrawal

Faculty of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad, INDIA

ABSTRACT

With wide use of internet as a medium of communication, it is a need of time to have service oriented architecture. Web services are loosely coupled independent executable software components that help to provide data for communication to web applications. Data formats used in web services have pre defined data exchange formats like XML and JSON. These web services works on heterogeneous platforms having different network architecture and operating systems. The basic architecture extensively used in web services is RESTful architecture. The RESTful web services has REST APIs , that works on four simple methods of POST, PUT, GET, DELETE. This paper provides an insight of various features required to compare two widely used data exchange formats and present a comparative analysis on these features conducted by the researchers. This comparative analysis has proved that JSON has significant impact as compared to XML that provide its future scope for comparison in these two data sets on the basis of efforts required to learn and implement JSON objects.

KEY WORDS
Web services, RESTful architecture, data exchange formats, XML, JSON

Received: 14 Feb 2019
Accepted: 18 March 2019
Published: 20 March 2019

***Corresponding Author**
Email:
vishaw.fca@mriu.edu.in
Tel.: +91-9711135115

INTRODUCTION

Due to wide increase of Internet for communication among many applications, developers are thinking more to simplify the concept of communication between client and server. The term web services were coined. Web is the model of communication between client and server over a network. A network can be internet or intranet. Web allows performing all type of communication over the Internet with the help of HTTP. All the information over the web is in the form of documents or other web resources. To access these resources there is requirement of URL. All the elements on the web is linked with each other making its model architecture complicated. Evolution of web starts from static web pages to dynamic web pages, to interactive web pages and finally to reactive web pages. Reactive web pages not only communicate with the users but also adapt the screen resolution and screen size. Service term is used for any software function performed for any business query, accessing data in various format or too perform certain comprehensive function like authentication etc. A service can be used to share logical functions across different applications. A web service is an independent executable software web component which is usually responsible for performing certain services required for reusable business function. A web service is a web application component that uses standardized format like XML and JSON to interact with other web applications over internet. Earlier data processing activities were integrated with web application development and made it a tedious task to access everything on the same machine i.e. server. Gradually with the development of technology , data processing service were separated and then this process data is used with web applications at the front end to generate certain specific output as per user specification. So web services are independent data processing services that provide processed data in special data exchange format like XML and JSON and provide this data to web applications. Web services are processed on different servers and provide the response to the client. Web services doesn't required any specific platform and they are loosely coupled , that is the reason they are usable in various types of business applications and generalized system factions. Thus a web service doesn't require a user interface where a web application has a GUI or a user interface to interact. Web services allows communication between various applications to be used with over Internet, here no human interaction are required. We can access various web services by HTTP methods specified GET, PUT, POST, DELETE etc. for example Google maps are the web services that can be used by web applications to display maps. To get the response from these service web Applications sends co-ordinates and get responses. [Fig. 1] below illustrates web service architecture.

Basic architecture of web services are of two types SOAP and REST, where SOAP is a set of protocols and REST is a software oriented architecture having rest APIs. The SOAP is a XML based protocols that defines the format of messages for communication on web services enabled web applications. The heterogeneous environment over internet demands a standard data exchange format representation that is designed in XML for SOAP architecture. SOAP message structure has three basic components: The SOAP header, the SOAP body and the SOAP fault.

The SOAP header is used to pass the information related to application. Authentication credentials and definition of complex types are embedded in header. The SOAP body contains the actual detailed message. The SOAP fault is an optional element that defines the nature of error message that can occur while processing. SOA is essentially a collection of services that required communicating with each other. Basic SOA architecture always has service provider and service consumer. In normal RPC style of communication, client just generate a request by calling a method and server access these requests and send the desired responses to the client. During this process requests are send in a pre defined format and this encapsulation of requests in SOAP message is called marshalling. At the server end, these requests are unwrapped for the generation of appropriate responses is called demarshalling. REST architecture is a style of software architecture that does not have pre defined standards and protocols. These architectures, also known as RESTful architecture, has proved a wide choice for web services

implementation. RESTful architecture is uniquely characterized by state and the functionality is divided into distributed resources. RESTful architecture has four basic methods for data access over web services.

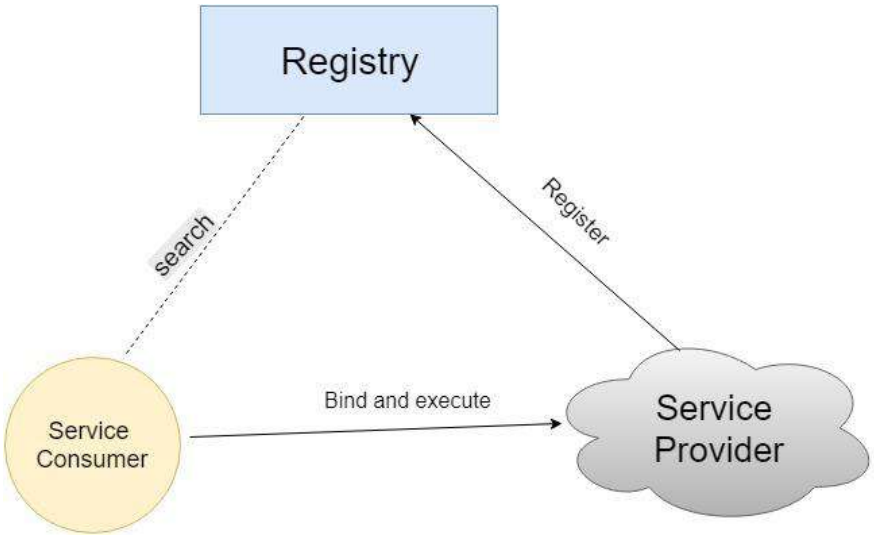

Fig. 1: Web Service architecture

LITERATURE REVIEW

Attempt has been made in the literature to find the use of various data exchange formats to improve the reliability of various standards of web technologies used in Internet. JSON and XML are 2 widely used data exchange formats having their own pros and cons. An attempt is made to discuss the contribution of some authors.

Research study [1] has given a broad outline of various features of JSON and XML as data exchange formats. And also describe the strength and weakness of both over each other. In this paper various criteria and their hierarchy for the comparison of JSON and XML formats is illustrated. A comparative analysis is performed and based on this, authors have concluded that XML bit edge over JSON and advised that a combination of can be used as per the requirement.

The comparative analysis in the paper [2] has given emphasis on selecting adequate data exchange format that can improve data transmission rate and performance significantly. The authors have shown comparative analysis between the two formats by considering 2 different scenarios on the size of java objects. The factors used for comparison were CPU time, memory utilization and data transferring time. The result in the study concluded that JSON is faster and used fewer resources.

The authors have proposed a web service model for data exchange between Heterogeneous databases. The data exchange format is in JSON format [3].

The benchmark comparison between the features of two most widely used data exchanged formats JSON and XML has provided a clear picture of performance of the two. On the basis of parameters stream size, memory foot prints and marshalling and unmarshalling time. JSON data format is best choice among the two [4].

In a paper [5] authors raised the issue of importance of data exchange formats in service based applications. Authors had used parameters like serialization and deserialization in various XML and JSON formats and concluded that JSON based data can be processed more efficiently.

In this paper [6] the authors presented the need of data storage on cloud and performed in-depth analysis of two most widely used data exchange formats JSON and XML based on a Web API implemented with PHP framework. Both formats were analyzed on two criteria, the response speed (in seconds) and the size of the data received (in KB). The overall evolution stated that JSON data exchange formats are better in terms of data size and speed over XML. The paper has provided with a future scope to convert XML data formats used in web services to JSON data formats.

In a paper authors emphasized the use JSON data format for storing the data in relational databases. The authors have used four most commonly used databases and represented their lesion with JSON to

represent database schemas. This is possible due the property of JSON to represent objects into key-pair values. An analysis was performed by testing various queries with this model representation. This provides a scope for JSON data formats to be used as schema representations of relational databases [7].

In another paper [8] the authors have propose a formal data model for JSON documents and, based on the common features present in available systems using JSON, the authors define a lightweight query language allowing to navigate through JSON documents. This proposed model has a prospect to provide theoretical framework for JSON documents.

In other study [9] the authors have raised the need of semantic web, as it allows easy retrieval and accessing of information by both man and machine. They have proposed a JSON data model to represent semantic web and linked data. The proposed model is PROV-JSON. This model helps in Provenance assertions.

PREVIEW OF XML AND JSON

XML stands for extensible markup language, a self descriptive hardware software independent tool for data exchange over Internet. It can be used with service oriented web application. Though data is encapsulated in tags but there is no pre defined mechanism for generation of tag definition. Tags can be created by the developers as per need and convenience. For example creating an object and encapsulating it in XML tags is shown below.

```
<MyName>
<first>Anu</first>
<middle>rag</ middle >
<last>Basu</last>
</ MyName >
```

All XML structures are self defined and well understood. The basic principle for effective design in XML is based on openness, extensibility, simplicity and scalability. XML has very simple and have logical syntax rules. The great support and availability of its understanding through tutorials has made its immense use in web, e business and portable applications.

JSON is java script object notation, has no predefined rules for organizing information. In JSON we can access logical data in human readable form. An example of creating a JSON data object in logical manner is as follows.

```
{
        "firstname" : “Anu”,
        “middlename”:”rag”
        "lastname" : "Basu"
}
```

JSON instances are fully immutable. It means without call or with call its instances never changed its existing architecture but create new instances as and when required. In JSON pairing is used for key and value and always focuses on content to be exchanged then format of the content. This feature of JSON improves its data transmission speed and save the time. JSON is not bounded by formatting rules, provides remarkable reduction in the size of JSON data format.

Table 1: XML stability over JSON

XML	JSON
XML has a predefined structure and certain rules for document specification that can act as a base for grammar rules.	JSON does not have a pre defined rules and grammar to represent data formats. Thus require interfaces for intended results.
It provides Namespace conventions to avoid duplicate name conflicts and has a great support for extensibility.	Namespace methods are not implemented with JSON and hence reduced its strength being extensible.
XML is decayed long standard used in industry and act as a support for various frameworks and browsers.	Framework support is very limited for JSON as it is relatively new in industry.
Human readability of XML documents is high.	JSON documents looks highly cryptic, as there is extensive use of parenthesis delimiters.
XML is extensively used in product development having web services oriented architecture.	JSON has limited support for web services.

Table 2: JSON stability over XML

JSON	XML
JSON meta data requirement is very less, hence provides high bandwidth performance.	XML tags are converted meta data that act as an overhead and provide slow bandwidth performance.
JSON has inbuilt object notation supported by java script and hence easier to map into objects.	It is difficult to map XML data formats into its corresponding objects due to their tree structure representation.
Serializing and deserializing can be effectively implemented with little code.	Special java scripts are written for serialization and deserialization.
JSON use key value pairing for representation of data objects and hence a comprehensive formats.	Tag representation of data objects provides prolonged data exchange formats.
JSON data objects can be effectively implemented with AJAX based web applications.	AJAX toolkits don't have strong support for handling XML.

COMPARATIVE ANALYSIS OF JSON AND XML DATA FORMATS

This paper [2] has presented a research work which compare JSON and XML data formats. The authors presented two different scenarios, one with large data and second with different size of data. The comparative analysis of their transmission time for data in the form of XML and JSON is tabulated in [Table 1] and [Table 2] below. The conclusion driven from the statistics available shown JSON has overpowered XML.

Table 1: Context 1 XML vs. JSON Timing

	JSON	XML
Number Of objects	100000	100000
Total Time taken (ms)	7827.79	454469.47
Avg. Time (ms)	0.08	4.54

Table 2: Context 2 XML Vs JSON timing

	JSON	XML
Exp 1. Number Of Objects	10000	10000
Exp 1 Total Time (ms)	1106.57	30666.84
Exp 1 Avg Time (ms)	0.11	3.07
Exp 2 Number Of Objects	20000	20000
Exp 2 Total Time (ms)	1563.99	61927.29
Exp 2 Avg Time (ms)	0.08	3.10
Exp 3 Number Of Objects	30000	30000
Exp 3 Total Time (ms)	2276.19	92968.13
Exp 3 Avg Time (ms)	0.08	3.10
Exp 4 Number Of Objects	40000	40000
Exp 4 Total Time (ms)	3003.36	123819.91
Exp 4 Avg Time (ms)	0.08	3.10
Exp 5 Number Of Objects	50000	50000
Exp 5 Total Time (ms)	3748.68	155008.74
Exp 5 Avg Time (ms)	0.07	3.10

CONCLUSIONS

With the increase in web based applications for data communications, has evolved web services utilization and provide researchers a new area of interest. Effectiveness of web service based data communication depends upon the selection of appropriate data exchange formats according to the need of the system. In

IIOAB JOURNAL
ISSN: 0976-3104

this paper Author has shown researchers' work to show significant impact of JSON over XML in the scenario where the data transmission speeds is a considerable issue. JSON as a better data exchange format is a widely accepted fact.

Not much of the work has been carried out by the researchers to authenticate the migration of development of web services from XML-SOAP architecture to RESTful-JSON architecture. This provides a future scope to carry out the impact of JSON data exchange format in parsing and efforts required in learning.

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

REFERENCES

[1] Haq Z, Khan G, Hussain T. [2014] A Comprehensive analysis of XML and JSON web technologies, New Developments in Circuits, Systems, Signal Processing, Communications and Computers.102-109.
[2] Nurseitov N, Paulson M, Reynolds R, Izurieta C. [2009] Comparison of JSON and XML Data Interchange Formats: A Case Study, 22nd International Conference on Computer Applications in Industry and Engineering.
[3] Datt R. [2015] Data Exchange Model Using Web Service For Heterogeneous Databases, International Journal of Advanced Research in Engineering and Technology, 6(4): 107-111.
[4] Zunke S, D'Souza V. [2014] JSON vs XML: A Comparative Performance Analysis of Data Exchange Formats, International Journal of Computer Science and Network, 3(4): 257-261.
[5] Peng D, CAO L, XU W. [2011] Using JSON for Data Exchanging in Web Service Applications, Journal of Computational Information Systems, 7:5883-5890.
[6] Breje A, Gyorodi R, Gyorodi C, Zmaranda D, Pecherle G.[2018] Comparative Study of Data Sending Methods for XML and JSON Models, International Journal of Advanced Computer Science and Applications, 9(12):198-204.
[7] Piech M, Marcjan R. [2018] A new approach to storing dynamic data in relational databases using JSON, Computer Science Journal, 20: 1-18.
[8] Bourhis R, Reutter JL, Suarez F, Vrgoc D.[2017] JSON: data model, query languages and schema specification, Cornell University Computer Science Journal, 1. DOI: http://dx.doi.org/10.1145/3034786.3056120
[9] Pandey M, Pandey R. [2017] JSON and its use in Semantic Web, linternational Journal of Computer Applications,164(11): 10-16.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

DEBUGGING MICROSERVICES WITH PYTHON

Sameer Shukla*
System Architect, HCL America, Fort Worth Texas, USA

ABSTRACT

The modern world applications are expected to be highly available, resilient and the applications are supposed to deal with many concurrent requests and a huge data set (Terabytes) now is a reality. To deal with such demanding requests a new architecture style has been introduced known as Microservices architecture where the services are broken into many small services, where each service is designed on the Single Responsibility Principle. There are situations where there are more than 100 microservices are running in parallel the major challenge comes in front of developers is how to test or debug all these microservices before moving to production and analyzing data during their testing and discovering important information, conclusions and taking corrective measures. In this paper I am trying to showcase an approach of testing, analyzing and debugging data in the Microservices architecture using Python and some of its fantastic packages from Developers Standpoint. I will highlight some of the important issues Developers and Testers face on daily basis in their testing because of the complexities Microservices architecture brings with itself and how we can use these Python packages to solve those issues, irrespective of the language in which the services are developed it can be Java, Scala, .NET or any other language.

KEY WORDS
Microservices, Python, Kafka-Python, Pandas, SQLAlchemy

INTRODUCTION

Microservice is SOA in new cloth, in Microservices architecture complex application is divided into various small services which can be developed and Scaled independently, and they can communicate with each other which indicates that we are developing a distributed system. Microservices are Fault Tolerant in nature meaning if one service is down others will keep running fine and the failed service has no impact on others. Another very useful advantage of Microservices are they can be Polyglot in nature consider a scenario where few Microservices needs to deal with lots of records (in Petabytes) here we can use NoSQL say Cassandra as our database and few microservices needs to deal with only a few millions of records here we can very well use traditional RDBMS as our data store. Working in Microservices architecture brings Clarify of Domain for the entire team since the application is composed of several services it's easier to understand the functionality of every service in the Microservice Architecture. Microservices have few disadvantages as well like Developing Distributed systems can be very complex and Testing of Microservices is a very cumbersome process, in Monolithic architecture performing Testing was very easy as there is one huge service which was doing everything but in Microservices architecture there can be hundreds of tiny services you can imagine Testing of such hundreds of tiny services is very challenging. Similarly debugging of these tiny services is a very cumbersome process for developers, this article focuses on a unique approach of debugging Microservices using Python [1] and its lightweight packages. No matter in which technologies you have developed your services, but we can very well use Python [1] and packages for debugging.

Received: 6 Mar 2019
Accepted: 7 May 2019
Published: 17 May 2019

Challenges in microservices architecture

No Technology or Architecture styles are perfect, similar is the case with Microservices it brings certain major challenges with itself. Few of them are Testing Microservices, Monitoring Microservices, Debugging Microservices, Learning Curve of the Technologies because of Polyglot Architecture, Integration Pain Points

Let's investigate more detail what exactly the challenges are in Testing and Debugging in Microservices. First, we will understand the Use Case and the challenges within that Use Case, and we will look at the approach to overcome the Challenge.

Use case

Let's discuss a real-world scenario which can be divided from monolith to actual functional services, Reliant Energy or Salt River Project (SRP) Energy would be an ideal microservice architecture use case.

Fig-1 shows the typical microservice environment where each service is responsible for one use case based on the Single Responsibility Principle. For ex: Customer Profile Service is responsible for managing customer data. Billing Service is responsible for handling customer billing. Consider a scenario where you are developing a Moving Microservice which is dependent on Consumption Microservice for its data, now something is wrong with Consumption Microservice data which is developed by some other developer or team then the most challenging part is to identify the problem in Consumption service. Let's look into it in detail [Fig-1].

Corresponding Author
Email:
sameer.shukla@gmail.com
Tel.: +1-480-754-9793

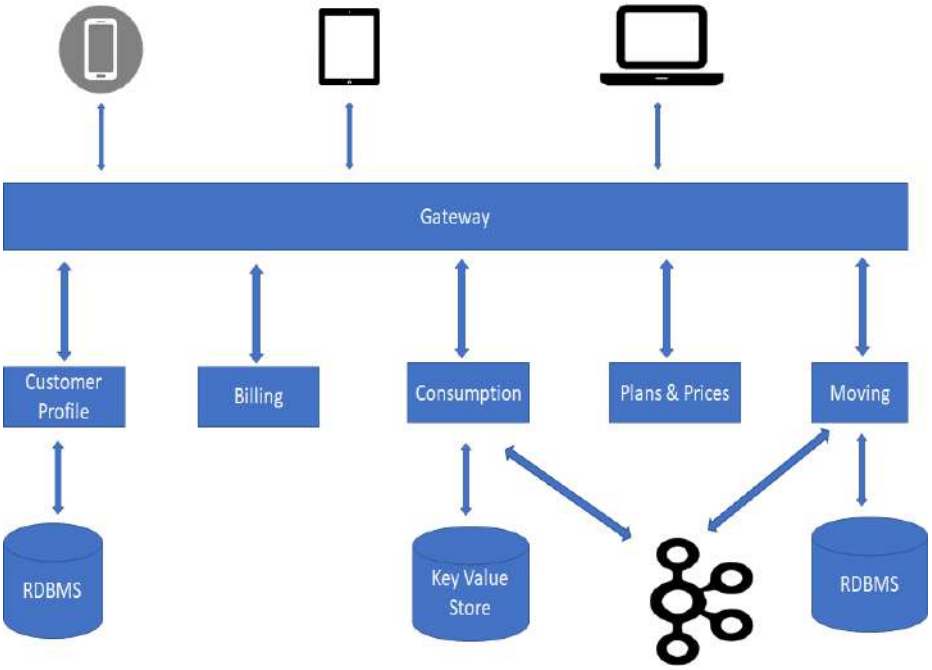

Fig. 1: Microservices environment

ISSUES IN TESTING/DEBUGGING MICROSERVICES

Composite nature

It is recommended that Microservices should be of Atomic Nature, meaning each service should have its own database otherwise if services share a common database than they are the candidate of Single Point of Failure if the common database is down all our services are down. Instead, if each service shares a separate database and then if one of the databases is down it has no impact on other services and they continue to work fine. For example, in our use case if "Customer Profile" service database is down it has no impact on "Billing" Microservice.

So far so good, services will work fine independently as they are loosely coupled but imagine a scenario where there are 100 microservices are running in our system and we want to test the entire pipeline, it will not be easy to test the pipeline and bigger issue is the data verification. Sometimes services transform the data based on the event received and we want to test our data at every step.

Another scenario, if something went wrong in Production, say service is expected to return some data but it's not as per the expectation how we will verify which service is the culprit because each service is atomic.

Logging

Logging and Monitoring are not easy in Microservices, many microservices each one of them logging and reporting issues together it will be very challenging to go and checks logs in individual services and finding the service and the root cause of the error. The solution to this problem is to externalize logging by pushing log messages to Kafka Topic and later analyzing them again with the help of a few tools like Kibana via Elastic Searches.

Testing dependent services

Again, the issue is same if services are dependent on each other say one asynchronous service needs to invoke a RESTful service to validate the data before saving, update or deletion testing of such dependent services and verifying the data the dependent service is returning is little difficult.

Testing Actor models

If we are developing reactive microservices, Akka [3] is the default choice for building reactive microservices. But how can we test actors which are dependent on other services and actors?

Verifying data in key-value store

As discussed in the Use-Case we are using Polyglot architecture in our system and one of the services is using a Key-Value store for persisting and retrieving data say, Cassandra [12], now Cassandra works on Partitioning key and Clustering key mechanism but during our development if we want to verify whether some other column exists in Cassandra database or not but we don't have a Partitioning key, how we should handle such problems during our development phase

IIOAB JOURNAL
ISSN: 0976-3104

Tools for debugging and testing

We are going to use the following Python [1] Packages for our debugging and analysis

- Pandas [9]
- SQLAlchemy [2]
- Requests module [10]
- Kafka-Faust [5] / Kafka-Python [7, 8]
- PySolr [11]
- PyAkka [6]

The beauty in these packages is they are lightweight, easy to learn & use, tools can be developed using these packages in very less time. A lot of online help on these packages are available and their documentation is also very crisp

METHODS

Debugging and analyzing RDBMS

Scenario: I am a developer who is working on the development of some Microservice but for some reason, I need to check the "Customer Profile" service and its data for debugging purposes and analysis reasons. Two things a developer can do here either Invoke a Customer Profile service or directly dip into a database and do the verification. I have decided to directly go into a database, Firing manual queries one by one will be a tedious and time-consuming job as I need to check many things like whether the customer is an active customer or not, what are the details of the customer based on Customer Id for billing related purposes, since how long the customer is an active customer to recommend him some better plans etc. Here I would like to use a package called SQLAlchemy [2] and combine it with Pandas [9].

SQLAlchemy

It's a lightweight database toolkit written in Python, the advantage of using SQLAlchemy [2] toolkit here is we can very easily use it in Jupyter notebook, and we don't have to create the entire framework or write hundreds of lines of code just to communicate to a database, hardly in 7 lines we can query any RDBMS and get result. Another powerful advantage of using SQLAlchemy [2] is that we can define the database schema in the application code itself, but we are more interested in only reading the data from the database. The code is very readable because it is written completely in Python. Our scope here is limited that's why I am not going into the details of ORM here].

Pandas

Pandas [9] is an excellent data analysis library, it can load data from any data source such as CSV, JSON, database and creates a Python object called DataFrame that looks very similar to the table. Once DataFrame is created we can perform a variety of operations like Grouping, Filtering, sorting by keys and values, Creating Series from DataFrame, etc.

CREATE TABLE db. customer_profile (customer_id INT, first_name VARCHAR (20), last_name VARCHAR (20), email VARCHAR (20), is_active boolean, created_date DATE) [Fig-2].

```
In [2]: import pandas as pd
        import pymysql
        from sqlalchemy import create_engine
        pymysql.install_as_MySQLdb()
        engine = create_engine("mysql://root:root@localhost/db")
        conn = engine.connect()
        result = conn.execute("SELECT * from customer_profile").fetchall()
        result[:2]

Out[2]: [(100, 'Sameer', 'Shukla', 'sshukla@gmail.com', 1, datetime.date(2019, 2, 28)),
         (101, 'John', 'Doe', 'jd@gmail.com', 1, datetime.date(2019, 2, 28))]

In [3]: df = pd.DataFrame(result)
        df.columns = result[0].keys()

In [16]: # Find all inActive Customers
         df[df['is_active'] == 0]

Out[16]:
```

	customer_id	first_name	last_name	email	is_active	created_date
3	103	Dav	P	dp@gmail.com	0	2019-02-28

0 indicates Inactive Users and 1 Indicates Active Users.

Fig. 2: Example of Pandas with SqlAlchemy

We are simply fetching the result from the Database using SQLAlchemy [2] and giving the result set to Pandas [9] which then created the DataFrame object on which we are executing our analysis. Here we are looking into all the Inactive Users in the Customer Profile Database

Debugging and analyzing rest service

I am extending the scenario mentioned above, now I have decided to invoke the RESTful service and then validate the JSON Response. The package I will like to prefer here is "requests" combining it with Pandas.

Requests package

Requests [10] is a simple, lightweight HTTP library or we can say a simple HTTP Client library. Advantage of using this library is the same we don't have to create the entire framework just for invoking a RESTful service, plus combining this library with Pandas [9] makes it more powerful because sometimes a JSON response can be huge thousands of line, try invoking https://api.github.com/events [4] service you can see response is of approximate 2000 lines, identifying or locating particular field, in this case, would be a difficult task. The approach here is simple, invoke the service and give the JSON response to Pandas and perform operations on DataFrame. For the above URL counting how many Push Events or Create Events the response has is impossible to achieve, but with Pandas, it's just a line of code, have a look [Fig-3].

```
In [2]: import requests

In [3]: # Invoke Service
        request = requests.get('https://api.github.com/events')

In [26]: # Pandas
         import pandas as pd
         import json
         df = pd.DataFrame.from_records(request.json())
         # Counting Push events
         df['type'].value_counts()

Out[26]: PushEvent           16
         CreateEvent          4
         WatchEvent           4
         ForkEvent            2
         PullRequestEvent     2
         GollumEvent          1
         IssuesEvent          1
         Name: type, dtype: int64
```

Fig. 3: Example of Pandas with requests package

After executing the Rest service, we are taking the JSON response and handing it over to Pandas, the DataFrame is returned by Pandas on which we have executed our analysis which is count the Push Events

Debugging and analyzing Kafka

There are two very important packages are available to deal with Kafka, first is Kafka-Python and another one is Kafka-Faust. Faust is a Stream Processing Library; we can process both Streams and Events using this library. Another one is Kafka-Python, it's a Python Client for Apache Kafka, using Kafka-Python library we can consume messages directly from Kafka by subscribing to Kafka Topic and we can also produce messages to Kafka Topic. To test the Moving service, which is consuming messages from Kafka Topic, we can write a lightweight Kafka Producer using Kafka-Python library and post a message on the topic from there Moving service will pick up the message and further process it. We enhance our Test client using SQLAlchemy + Pandas and analyze the data in the RDBMS [Fig-4].

Other important packages

PySolr: It's a lightweight wrapper for Apache Solr, you can query Solr server and analyze the data returned based on the query executed. Use Case for PySolr can be, redirect message from Logging Kafka Topic and index the logs using Apache Solr and query the Solr Server using PySolr.

PyAkka: It's a Python Implementation for Actor Model [3].

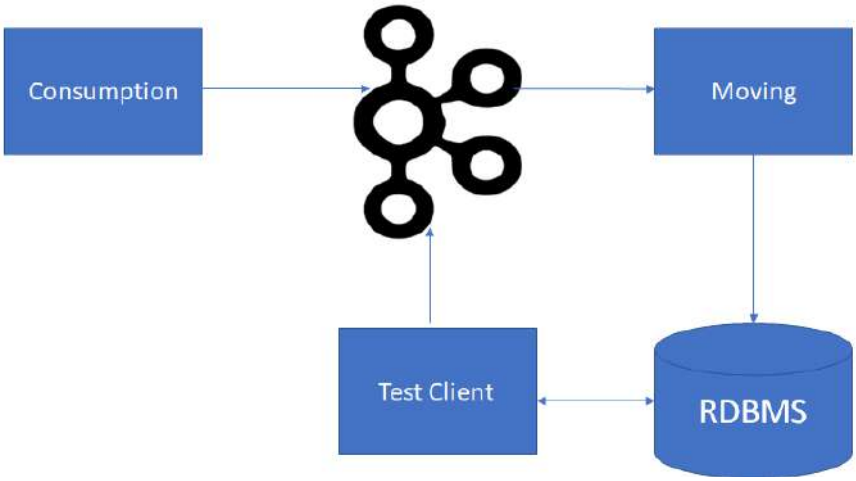

Fig. 4: Kafka in microservices

RESULTS AND DISCUSSION

In the modern era Gigabytes, Petabytes of data is the reality and distributed system and reactive systems are the need of the hour. Engineers must bring unique approaches to every aspect of SDLC for being efficient. In hundreds of Microservices environment it's difficult for an engineer to identify the issues in the system, Python is not purely functional but it's tools are amazing, like we don't have to deploy an application or we don't even need an IDE if we want to debug something a simple Jupyter Notebook is self-sufficient, on which we can do almost everything, all the packages I have mentioned in this article can very well run in Jupyter notebook, we can immediately verify what our microservices has done or we can quickly check the dependent services issues or data. I strongly recommend to the engineers to use Python [1] a lot in whatever environment they are working and whatever tools or technologies they are using.

Debugging and Testing Microservices with Python [1] and related packages in a unique approach. Each package I mentioned in the article is lightweight, for example using SQLAlchemy [2] for database operations has several benefits like we can define the database schema in the application code itself, we need to construct database queries in Python which brings FRM (Functional Relational Mapping) like flavor with itself. Requests package is very convenient to use we don't have to write hundreds of lines of code just to invoke a service the Requests package do it for us. The combination of these packages with Python Pandas [9] brings analytics also into the picture, Pandas is a package which is widely used in Data Analytics world and using Pandas with SQLAlchemy results can do wonders for us. Similarly, there are various other packages like Matplotlib for graphs which we can be very well used in the Microservices world, giving Pictures (Graphs) like Pie chart, Bar Graphs, Histograms to our data can be very useful for everybody as picture speaks a thousand words, say if we want to show the top 10 or last 10 kind of results from our Database how convenient and useful it is when we show them using bar graphs rather than seeing the plain data. Another Advantage of using Matplotlib + Pandas + SQLAlchemy or Requests or PySolr [11] can help us in identifying the problems in our system say we are using Cassandra as our data store and if we want to see whether partitioning balancing is proper or not in our system these graphs can help us in identifying such issues.

CONCLUSIONS

The Microservices architecture in today's world is a reality, there are many tools, languages, frameworks are available on which these Microservices are developed. We need to be smart and quick in figuring out the issues these services can have in Development or Production Environments, otherwise it will defeat the purpose and advantages this Architecture style brings with itself. We need to know all the Integration Points and we should be having some tools with us to figure out the problems in those Integration Points or in the Services. Python [1] and its packages provide us the best options in front of us to identify such problems. As an engineer in today's worlds, it is mandatory that we should not restrict ourselves to one technology or framework, this Python [1] approach has helped me big time in resolving the issues I have mentioned in this Paper. I strongly recommend engineers to follow this approach in quickly identifying and fixing issues in Microservices environment

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] http://docs.python-requests.org/en/master/
[2] https://www.sqlalchemy.org/
[3] https://doc.akka.io/docs/akka/2.5/testing.html
[4] https://api.github.com/events
[5] https://faust.readthedocs.io/en/latest/
[6] https://www.pykka.org/en/latest/
[7] https://kafka-python.readthedocs.io/en/master/usage.html
[8] https://www.confluent.io/
[9] https://pandas.pydata.org/
[10] https://pypi.org/project/requests/
[11] https://pypi.org/project/pysolr/
[12] http://cassandra.apache.org/

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

INTERNET OF THINGS (IOT) BASED TRAFFIC INFORMATION SYSTEM

Robin Swamy[1], Vidushi Mangal[1], Anupriya Jain[1]*, Sonia Duggal[1], Prasenjit Banerjee[2]

[1]Dept of Computer Application, Manav Rachna International Institute of Research and Studies, Faridabad, INDIA

[2]Accendere Knowledge Management Services, CL Educate Ltd, New Delhi, India

ABSTRACT

One of the catch phrase in the Information Technology these days is Internet of Things (IoT). The future of technology can be termed as Internet of Things, which will convert the real world objects into smart virtual objects. The intent of IoT is to unite everything in our world under a widespread infrastructure, giving us not only control of things surrounding, but also update us about the state of the things. In the light of all these observations this paper provides a general idea of (IoT) by highlighting on traffic control system. The current revolution in Internet, mobile industry, and machine-to-machine (M2M) technologies could be seen as the beginning phase of the IoT. IoT is expected to unite different technologies to enable new applications by bridging physical objects together in support of intelligent decision making. IoT is facilitated by the latest developments in radio frequency identification (RFID), intelligent sensors, communication technologies, and Internet protocols. The basic objective is to have traffic sensitization directly without human intervention. This paper will propose a method for traffic management system which can ease out the human being's life. This paper starts by the introduction to the IoT. Then, we give an overview of some technical problems of current traffic system which is followed by the solution to the current system using IOT. Finally it gives some future recommendation for effective implementation.

KEY WORDS
IoT, RFID, Meta Data, GPS

INTRODUCTION

There could be no perfect definition to describe IoT however it has set new standards to revolutionize the human life. The various groups customize the definition of IoT as per their requirements. "Kevin Ashton is known first for coining the term the Internet of Things". IoT can be applied whenever there is network connectivity to practical world via multiple sensors. Few other definitions are given as:

IOT can be termed as "Meta-Data" where Meta word could be replaced by smart machine. It means is a combination of various intelligent machines which are based on each other while the time comes to collect the data or gathering of the information [1].

"An open and comprehensive network of intelligent objects that have the capacity to auto-organize, share information, data and resources, reacting and acting in face of situations and changes in the environment" [2].

It is a method which uses communication technology through smart sensing via big data to provide a complete package for a product. In this way it provides better control without human intervention which can be applied in any application areas. It can suit in any kind of environment. They enhance the collection of data, methods of automation, diverse operations, and much more with the help of smart devices and powerful technologies. IOT enhances the possibilities to apply in different walks of life and also improves its efficiency. IoT is an extended version of the previous and the new generation technologies for sensors, networks and artificial intelligence.

Received: 29 Mar 2019
Accepted: 10 May 2019
Published: 26 May 2019

IoT accomplish latest innovations in technology. It also helps in reducing hardware prices as represented in Fig. 1. The best output of IoT is that it revolutionized the way people think about technology. This modern and emerging approach brings considerable development in the end release of services. It impacts people in social as well as economic manner. Even we could see the political impact of these technological changes as depicted in Fig. 2.

KEY FEATURES OF INTERNET OF THINGS

The term "Internet of Things" abbreviated as IoT is derived from two different words "Internet" and "Things". In short in simple way it can be defined how internet can be applied in practical things. Internet is a superset of networks where every computer system in the world can be connected in communicate through networks. The key features of IoT may include artificial intelligence, connectivity among machines, communication networks, sensors, active engagement, and the use of small devices. They could be better explained through some definitions as given below:

Artificial intelligence- IoT essentially makes everything smart virtually. IOT make use of every area of life with the power of diverse data collection, algorithms of artificial intelligence, and data communication networks. Data could be as simple as the functioning of car seat belt or as complex as analysis of the data generated by the airplanes.

***Corresponding Author**
Email:
anupriya.fca@mriu.edu.in
Tel.: +91 9911293897

IIOAB JOURNAL
ISSN: 0976-3104

Communication network- Communication networks are the platform of up-to-date automation systems. It is a technology that refers to exchanging of data between diverse smart devices within a system. Communication networks provide a number of advantages in terms of local and global connectivity.

Connectivity- It means various different types of machines are connected together to works as a single device. The networking now means they are not reserved for major ISPs whereas they are diversified to every nick and corners in the world. IoT creates large networks through the connectivity of various small networks through the networking devices.

Sensors- Sensors are used to sense a situation which is not behaving normal for example suddenly the sound pollution on a road increases from 5 percent to 50 percent. It then reports it to the other machines connected to it. The sensors are integral and vital component if IoT. The sensors convert a static network in to a smart network which has the capability of applying in to the live scenarios.

Fig. 1: Representation of Internet of Things [3].

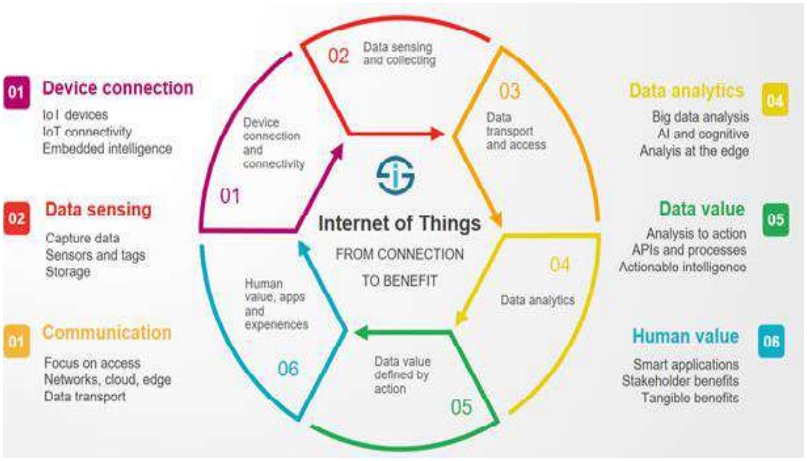

Fig. 2: From connecting devices to human value [4].

Real world interaction- In the conventional methods the communications with networks occurs only through static mode. But in case of IOT every machine connected to each other search for an abnormal condition surrounding it. And for this reason all connected machine has to be active all the time. IoT introduces all together a new concept through which we can have real time interaction in terms of content, product and services.

Miniscule components- With the emergent technologies the innovative devices have reduced size, cost and increased efficiency. The motive of IoT is to create small devices to for one or more purposes which can deliver its precision, scalability, and adaptability.

Advantages of IoT: IoT has affected the life of human in all aspects. The best part of IoT is that it can be applied in a very simple or in a complex situation of real world. Some of the advantages that IoT are as follows-

IIOAB JOURNAL
ISSN: 0976-3104

Improvement in customer engagement: Current analysis of data suffers from hidden areas and considerable loss in accuracy; and as explained, engagement remains inactive. IoT works in such a way that it provides better and more effective solutions.

Smart technology utilization: Better you know, better you use. If you have experience of a technology you can utilize it up to great extent. IoT helps in unfolding the use of diverse devices in the different areas of Human life.

Effective management: IoT gives clear picture of areas which needs or have a scope of improvement. Current technologies give us surface data, but IoT provides deep data in all the directions which leads to better management of resources.

Big data analysis: Conventional approach has the drawbacks that it is designed only for the passive use. IoT utilizes the big data and apply it in the day to day life and gives the transparency and understanding of its functioning.

Limitations of iot: As compare to limitations IoT has lot many advantages. IoT faced a number of depicted challenges which are listed below:

Security: IoT forms network of different types of devices connecting with each other. Because of it the system offers comparatively less control despite of different security measures. This could make a room for various kind of violation of security and privacy.

Privacy: There is a flow of all kind of user/device information without the active participation of user which can breach the privacy of user.

Complexity: IoT connects different type of hardware and algorithms to connect different type of machines so it becomes difficult to outline, and deploy. Even its maintenance becomes a cumbersome task as it includes devices of diverse category.

Flexibility: Due to the complex nature of hardware and algorithms used in IOT flexibility becomes as major concern of IoT. It is difficult to find a device with several conflicting or locked systems.

Compliance: As it applies in any real world situation, it has to follow certain set of regulations which pose a challenge to it.

Hardware - The component used in IoT comprises of devices for a remote system for controlling the system. It includes devices like routers, bridge and sensors to connect the different network. The cost of maintaining all these devices is quite high.

LITERATURE SURVEY

There were a number of researchers who worked with the traffic issues and its controlling. As a result of their hard work a number of different approaches have been invented. Pang et al. [3] proposed a traffic flow prediction mechanism based on a fuzzy neural network model in chaotic traffic flow time series. Bhadra et al. [12] applied agent-based fuzzy logic technology for traffic control situations involving multiple approaches and vehicle movements. In [13] the authors developed strategies to integrate different dynamic data into Intelligent Transportation Systems. Patrik et al. [14] proposed a service-oriented architecture (SOA) for an effective integration of IoT in enterprise services. Singh et al. [15] presents a discussion on Internet oriented applications, services, visual aspect and challenges for Internet of things using RFID, 6lowpan and sensor networks. Gourav Misra et al. [16] Mentioned the visions, concepts, technologies, various challenges, some innovation directions, and various applications of Internet of Things.

PROBLEM DEFINITION

Struggling with the cost of rupee, inflation and political havoc, there seems to be no narrowing of crises in India. According to a recent study conducted by the Transport Corporation of India and IIM (Kolkata), “India faces a loss of Rs 600bn ($10.8bn) a year” [10] due to congestion, slow speed of freight and waiting time at toll plazas. Toronto is facing a similar situation where $3.3 billion are lost every year. A report published by the Organization for Economic Co-operation and Development (OECD) in 2009 revealed that billions of dollars are lost every year [4].

The continuous boost in the traffic congestion level on public roads, especially at busy hours, becomes a serious issue in the world. It has become a foremost concern for the traffic management analyst and the decision makers. The conventional approach for traffic management, control and surveillance are not equipped in terms of its efficiency, economy, and the efforts needed for the implementation. India is the second most populous country in the World and is a fast growing economy as well. It does have dreadful

IIOAB JOURNAL
ISSN: 0976-3104

road congestion problems in its urban areas and rural areas as well as they are also growing at a considerable pace. The rate at which Infrastructure grows is comparatively inadequate as compared to the increasing number of automobiles on road, due to spatial and cost issues.

Moreover, Indian traffic is non-lane based, chaotic and not as discipline as in other countries. It requires a traffic control system, which is more phenomenal from the developed Countries. Intelligent and smart management of traffic flow can decrease the negative impact of traffic congestion at a considerable rate.

Traffic congestion is at its alarming situation in developing Countries like India. Growth in urban and rural population is contributing significantly to the increasing number of automobiles in the cities and villages. Congestion on roads results in slow moving traffic, which increases the time of travel. This gridlock can have a tremendous effect on different perspectives of our life. It can lead to road rage, wastage of fuel, increase in air pollution, impact on our personal life, career, future and even our safety.

EXISTING SYSTEM

Traffic congestion is one of the major troubles in metro cities; it not only leads people to inconvenience travel, but also spoil the urban environment. The urban road traffic control system usually includes signal control machines, traffic lights, Variable Message Signs (VMS) and other detectors.

In last few years, wireless networks are extensively used in the road transport management as they provide better cost effective options. Technologies like ZigBee, Bluetooth, Wi-Fi, (Radio Frequency Identification) RFID and Global System for Mobile communications (GSM) can be used in traffic control to provide money spinning solutions but with the ever increasing speed of vehicle on road we required a better solution for the problem. Moreover, these devices are not capable of handling the situation in case of traffic jams, gridlocks etc.

PROBLEM DEFINITION

Whenever people plan to go out, they prefer a route which is less congested. Getting stuck in a traffic jam is a nightmare. For this, people prefer to take the advantages of technologies like Google maps to update the information about the most preferred routes. But the basic limitation of Google maps is that they predict the traffic condition on the basis of number of android mobile phone not on the number of vehicles on the road. This information may or may not be convincing because there may be a situation that in a single multi wheeler, there are more than one passenger and each is carrying an android device or there may be a situation where there are a number of vehicles on the roads but no person is carrying the android mobile phone. In such scenarios, the output cannot match the real time scenario.

But getting the real time information of traffic situation is a dire need as it could have cascading effects on lives of human being. A person can miss a flight, train, and exam, interview etc. & the list is endless. The situation could be so dreadful that a person may lose his/her life.

PROPOSED SYSTEM

In this work the concept of IOT is applied in the transportation sector. Here the cameras are placed on the major traffic sensitive areas and the real time information of traffic is filtered. The abnormal traffic condition is stored in the cloud where it is extracted by the traffic control system (TCS). The traffic control system is responsible for disseminating the information to all the areas where this time bound information is required like hospitals, Airports, Railway station etc. which could further be used for the convenience of the human as shown in Fig. 3.

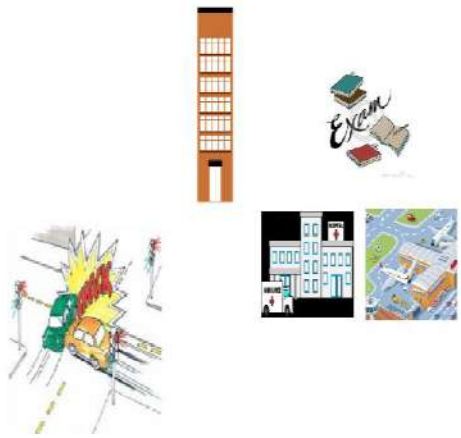

Fig. 3: Information passing through IoT connected devices.

FUTURE SCOPE

In the proposed system the current traffic information is propagated only to its respective traffic control system. So this method is restricted or limited only to its traffic control system what if, if the system fails. The future scope of this method should be to sensitize the traffic control system and also to broadcast the information to the entire location specific individual so that they can choose the alternate route. This system can also be improved if we restrict further vehicle to the traffic sensitive area. This IOT can also be linked to the satellite radio so as to transmit the information within real time.

CONCLUSION

The research was focused on reducing the traffic to ease out the life of human being. The method proposed by us has taken the advantage provided by IoT by giving real time data of traffic conditions of traffic sensitive areas. This method would cater to all the sections of the society satisfying their varying needs. IoT would take care of the real time condition which will outperform to the existing GPS, RFID systems. Consumer would benefit because it not only gives the actual traffic condition but it also facilitate the user by disseminating the information to all the areas where this time bound information is required like hospitals, Airports, Railway station etc. which could further be used for the convenience of the human.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Prasenjit Banerjee, Accendere Knowledge Management Services for his valuable guidance. However, the mistakes are all ours. It will not hamper his reputation.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Internet of Things (IoT Tutorial): Available at: https://www.tutorialspoint.com/internet_of_things/index.htm [Last accessed 22nd March 2019].

[2] Madakam S, Ramaswamy R, Tripathi S. [2015] Internet of Things (IoT): A Literature Review, Journal of Computer and Communications 3:164-173.

[3] Pang M, Zhao X. [2008] Traffic Flow Prediction of Chaos Time Series by Using Subtractive Clustering for Fuzzy Neural Network Modelling, Proceedings 2nd International Symposium Information Technology Application, Washington – DC. 23-27.

[4] What is the Internet of Things? Internet of Things definitions: Available at: https://www.i-scoop.eu/internet-of-things/ [Last accessed on 22nd March 2019].

[5] Internet of Things, Arshdeep Bagga,Vijay Medisetti Available at: https://books.google.co.in/books?id=JPKGBAAAQBAJ&lpg=PA1&pg=PA1#v=onepage&q&f=false {Last Access on 22nd July 2019].

[6] Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

[7] Intelligent Traffic Control System: Available at: https://electronicsmaker.com/intelligent-traffic-control-system [Last accessed on 22nd March 2019].

[8] How does Google Maps Show Traffic Updates: Available at: http://techwelkin.com/how-does-google-maps-show-traffic-updates [Last accessed on 22nd March 2019].

[9] COVER STORY: FABS IN THE INTERNET OF THINGS ERA: Available at: http://www.appliedmaterials.com/nanochip/nanochip-fab-solutions/december-2013/cover-story-fabs-in-the-internet-of-things-era [Last accessed on 22nd July 2019].

[10] India loses $10.8bn annually due to traffic congestion – study: Available at: https://arabiangazette.com/india-traffic-congestion-losses/ [Last accessed on 22nd March 2019]

[11] M. Pang and X. Zhao, —Traffic Flow Prediction of Chaos Time Series by Using Subtractive Clustering for Fuzzy Neural Network Modelling,” Proceedings 2nd International Symposium Information Technology Application, Washington – DC, 2008, pp. 23-27.

[12] Bhadra S, Kundu A, Guha SK. [2014] An Agent based Efficient Traffic Framework using Fuzzyǁ, Fourth International Conference on Advanced Computing & Communication Technologies.

[13] Katiyar V, Kumar P, Chand N. [2011] An Intelligent Transportation System Architecture using Wireless Sensor Networkǁ, International Journal Computer Applications. 14:22-26.

[14] Spiess P, Karnouskos S, Guinard D, Savio D, et al. [2009] SOA-based integration of the internet of things in enterprise servicesǁ, In: Proceedings of IEEE ICWS, Los Angeles.1–8.

[15] Dhananjay singh, Gaurav Tripathi, Antonio J, Jara. [2014] A survey of Internet-of-Things: Future Vision, Architecture, Challenges and Services, IEEE World Forum on Internet of Things (WF-IoT).

[16] Misra G, Kumar V, Agarwal A, Agarwal K. [2016] Internet of Things (IoT) – A Technological Analysis and Survey on Vision, Concepts, Challenges, Innovation Directions, Technologies, and Applications (An Upcoming or Future Generation Computer Communication System Technology, American Journal of Electrical and Electronic. 4(1):23-32.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

CLOUD COMPUTING SECURITY ISSUES

Shubhankar Solanki, Wasim Khan, Harshit Arora, Dr Anupriya Jain*
FCA Department, Manav Rachna International Institute of Research and Studies
Delhi Suraj Kund Road, Sector 43, Faridabad, Haryana 121004, INDIA

ABSTRACT

*Today, cloud computing is a trending way of computing in computer science. Cloud computing is a set of resources and services that are offered by the network or internet provided by the third-party, allowing sharing of resources and data among devices. It is broadly used in many organizations nowadays and becoming more sought after because it changed the way resources (IT) of an organization are utilized and managed. It provides lots of benefits such as simplicity and lower costs, scalable storage capacity, low maintenance, ease of use, backup and recovery, 24*7 availability, quality of service, automatic software integration, flexibility and reliability, easy access to information. While there is increasing use of cloud computing service in this new era, the security issues of the cloud computing are becoming challenging. Cloud computing must be more safe and secure to ensure the privacy of the users. This paper focuses on the most common security issues of using cloud and certain solutions to the security issues. Security is one of the most crucial aspect in cloud computing due to the sensitivity of user's data.*

KEY WORDS
Cloud Security, Service Models, Deployment models, Security Threats, Security Techniques

INTRODUCTION

Cloud Computing is a service model that produce various services in the form of on-demand services, it is accessible worldwide to everyone, everywhere and every time, including cloud referring to the web. In simple, Cloud Computing is a mixture of a technology that provides the hosting as well as storage service on the Internet. Its main intention is to provide scalable and affordable on-demand computing structure with superior quality of service levels [1]. Various national and international corporations are working on it and offer cloud computing services but they have not properly visualized the implications of accessing, processing and storing the data in a distributed shared environment. Many cloud-based application developers are struggling to include security. In multiple cases, the cloud developers simply cannot provide real security with the currently affordable technological capabilities [2]. Cloud computing concept is easy to understand as it allows us to share the resources on a greater scale .Distributed networks which requires less cost and is location independent. Resources on the cloud can be used by the consumers and deployed by the vendors such as Snapdeal, Google, IBM, Salesforce, Zoho, Rackspace, Flipkart etc[3]. Cloud computing model allows distributing the required on-demand services for various IT Industries. Benefits of Cloud computing are multifaceted. The most important benefit is that the users don't need to buy the resource from a third party vendor; rather they use the resources and pays for it as a service thus cloud helps the users to save time and also money. It is not only used by international companies but today it's also used by Small and medium enterprises [4].

Received: 21 Mar 2019
Accepted: 14 May 2019
Published: 2 June 2019

Cloud Computing architecture includes multiple cloud component which interacts with each other for various data which leads the user to access data on a faster rate. Cloud is mainly consists of front and the back end. Front end is the user side that is accessing the data, whereas the backend is the data storage device, server which makes the Cloud. Cloud computing is of three different categories viz, private cloud, public cloud and hybrid cloud. The private clouds are taken care by single organization and the public clouds are taken care on a larger scale. The Private clouds provide better security control and more flexibility than other cloud types. Hybrid clouds are the combination of Private clouds and Public Clouds that are used by various industries [5].

The benefits of cloud computing may be very appealing but nothing is perfect. Because the Cloud computing got many security issues especially on Data theft, Data loss and Privacy. This research paper lists the parameters that affect the security of the cloud, explores the cloud security issues and problems that the cloud computing service provider and also the cloud service customer face by such as loss of data, privacy, infected application and security issues [6].

The various Security issues of these systems and technologies are appropriate to cloud computing systems. For example, the network that interconnects the systems in a cloud computing has to be secured. Moreover, the virtualization paradigm in the cloud computing results the various security concerns. For example, mapping the virtual systems to the physical systems has to be carried out securely [7]. Data security includes encrypting the data as well as ensuring that the significant strategies are enforced for data sharing. Furthermore, the resource allocation and memory management algorithms has to be secured. Finally, data mining method may be applied to malware detection in cloud computing [8].

***Corresponding Author**
Email:
anupriya.fca@mriu.edu.in
Tel.: +91 9911293897

BACKGROUND REVIEW

Being the most trending technology of the age, the research is being done widely on Cloud Computing and especially on cloud security. In December 2008, Cloud Security Alliance (CSA) was formed with the aim to provide assured security within cloud computing environment. CSA launched "Security Guidance for

IIOAB JOURNAL
ISSN: 0976-3104

Critical Areas of Focus in Cloud Computing" as their initial product to help users get better insight about clouds and the security parameters [9]. The Cloud Computing Interoperability Group and the Multi-Agency Cloud Computing Forum have made lot of efforts to deliver efficient and effective controls to provide information security in Cloud environment Till date, many efforts have been made to find main security issues in cloud. It is described that privacy and the trust are the major security issues faced by the cloud computing. Security and privacy challenges to cloud computing are discussed in a detail which also addresses the security issue. It is claimed that cloud systems can't prosper without resolving security and privacy issues. A cloud computing framework and information asset classification model were proposed to help cloud users choosing different delivery services and models [2].

RELATED WORKS

The architecture of cloud composed of several service and deployment models.

Service Model

Software as a service (SaaS)

It is the top layer of cloud service model. The cloud service provider developed and hosts the software or application on the cloud infrastructure allowing the users to use it with various devices by using the thin client interface such as web browser. However the underlying cloud infrastructure, network, servers, operating systems or even individual application capabilities is not manageable by the users. It helps the users to save cost because licensing of the traditional packages is more expensive compared to the monthly fee for renting the application from cloud service [1].

Platform as a service (PaaS)

A middle layer of cloud service model that provides a software environment or platform for the users to design, develop, deploy and test their application without worrying about the underlying of the cloud infrastructure using the virtual servers of the cloud service provided. Therefore, the users can build their own applications running on the provider's infrastructure and they can have control over the deployed application they built [1].

Infrastructure as a Service (IaaS)

The user is allowed to rent the processing, storage and other fundamental computing resources to deploy and run arbitrary software which include operating system and applications .It provides basic storage and computing capabilities. It also has a data centre space that can help to handle workload [1].

Table 1. Comparisons of service model and examples

	SaaS Consume	PaaS Build	IaaS Host
Consumer	**End User**	**Application Owner**	**Application Owner**
Type of Service Provided	▪ Completed Applications	▪ Run Time scenario ▪ Cloud storage ▪ Integration, etc	▪ Cloud storage ▪ Visual server
Coverage at Service Level	▪ Application uptime ▪ Application performance	▪ Environment availability ▪ Environment performance ▪ No application coverage	▪ Virtual server availability ▪ Time to provision ▪ No platform or application coverage
Examples of Services Provided	▪ ▪ CRM ▪ E-mails ▪ Collaborative ▪ ERP	▪ Application development ▪ Decision support ▪ Web ▪ Streaming	▪ Caching ▪ Security ▪ Legacy ▪ System management

Deployment models

Public cloud

Third-party cloud provider owned a public cloud that is publicly accessible cloud environment. Any user can access it and they can store their data in the same cloud provided by the cloud service provider. The creation and on-going maintenance of the public cloud and its IT resources is managed by cloud service provider. In several scenarios the science of architectures explored in upcoming segments involve public clouds and the relationship between the producer and consumers of IT resources via public clouds [6].

Private Clouds

A private cloud is owned by a Single organization or users and it is not shared with the others. The user has physical control over the cloud infrastructure where everyone shares a common cloud infrastructure and it is more secure compared to the public cloud. The services provided by it is host services on private network that assist most corporate network and data administrators to become in-house service provider

IIOAB JOURNAL
ISSN: 0976-3104

efficiently. Same organization is technically both the cloud provider and cloud consumer in a private cloud [6].

Hybrid Clouds

A hybrid cloud is a cloud environment comprises of two or more different cloud deployment models. It is a combination of the public, the private or even the community cloud infrastructure which allows the transitive information exchange. It increases the flexibility of the cloud infrastructure where the users can implement the private cloud using the public cloud resources. For example, a cloud consumer may choose to deploy cloud services, process sensitive data to a private cloud and other, less sensitive cloud services to a public cloud. The result of this combination is a hybrid deployment model. Hybrid deployment architectures can be complex and challenging to create and maintain due to the potential disparity in cloud environments and the fact that management responsibilities are typically split between the private cloud provider and the public cloud provider organization[6].

Community cloud

The cloud infrastructure is shared among organizations that share the same concerns such as the mission, security requirement and policy. It may owned by more organization and it can exist on premises or even off-premises.

Each type of cloud model provides different level of control, flexibility and management. The users should choose the most suitable type of cloud computing model based on their own situation and their unique needs. This is very important since using inappropriate cloud model might cause the users to suffer for a great loss such as reduced organization efficiency and might suffer serious consequences like data breaches, data loss and corrupt data [6].

SECURITY ISSUES

Organization uses various cloud services as IaaS, PaaS, SaaS and the models like public, private, hybrid. These models and services have various cloud security issues. Each service model is associated with some issues. Security issues are considered in two views first in the view of service provider who insures that services provided by them should be secure and also manages the customer's identity management. Other view is customer view that ensures that service that they are using is securing enough [10].

Security issues

Data at rest is the major issues in cloud computing because users may store all their common, private, or even sensitive data in the cloud which can be accessed by anyone and anywhere. Data theft is a very common issue that are facing by the cloud service providers nowadays. Beside, some cloud service providers don't even provide their own server because of the cost effectiveness and flexibility [10]. There are also incidents like data loss which might be a serious problem for the users. For example, the server is suddenly shut down and causes data loss of the users. Furthermore, natural disaster might also cause data to be damaged or corrupted. Therefore, physical data location can be considered as one of the security issues in cloud computing [11].

Privacy issues

The cloud computing service provider must enforce their own policies to ensure the safety of the data stored by the users in their cloud model. The security of the data must be ensured and only the authorized person can maintain the cloud service model. The security of cloud computing should be done on the service provider side and also the user side. Cloud service provider should provide a good layer of security protection for the users where the users are not allowed to tamper with other user's data. The cloud computing is a good way to reduce the cost and provide more storage if and only if the security is ensured by both service provider and user [12]. The authors claimed that regulatory reform is essential to protect sensitive data in the cloud since one of the most challenging aspect in cloud computing is to ensure that the consumer have trust in privacy and security of their data [12].

Application issues

Monitoring and maintenance should be done by the cloud service provider frequently to ensure that the cloud is secure and not infected by the malicious code that have been uploaded to the cloud by the hackers or attackers with the purpose of stealing sensitive information or even damaging the information of certain users [14].

Threat Issues

There are lots of security issues regarding the cloud computing that have been widely used nowadays. There are top nine threat that pose severe danger to the cloud computing in year 2013 according to "The Notorious Nine: Cloud Computing Top Threat" by the Cloud Security Alliance (CSA). The top nine threat that have been mentioned in the white paper are:

IIOAB JOURNAL
ISSN: 0976-3104

Data Breaches

Data breaches are in all forms and have existed for years. Despite of new technology, Cloud computing and services still have data breaches. A study conducted by the Ponemon Institute entitled "Man in Cloud Attack" reports that over 50 percent of the IT and security professionals surveyed believed their organization's security measures to protect data on cloud services are low. This study used nine scenarios, where a data breach had occurred, to determine if that belief is a fact. After assessing each scenario, the report concluded that overall data breaches was three times more likely to occur for businesses that make use of the cloud than those that don't. The simple conclusion is that the cloud comes with a exclusive set of characteristics that make it more vulnerable [14].

Hijacking of Accounts

The growth and implementation of the cloud in many organizations has opened a whole new set of issues in account hijacking. Now attackers have the ability to use your login information to access your sensitive data stored on the cloud; besides, attackers can alter information through hijacked credentials. Scripting bugs and reused passwords are other method of hijacking, which leads attackers to easily and often steal credentials. Amazon faced a cross-site scripting bug that targeted customer credentials as well Phishing, key logging, and buffer overflow similar threats in April 2010 [13].

Insider Threat

An attack from inside your organization may seem impossible, but the insider threat does exist. Employees can use their authorized access to an organization's cloud based services to misuse information such as customer accounts, financial forms, and other sensitive information. And these insiders don't even need to have mischievous intentions. Imperva, inc., has published, "An inside Track on Insider Threats". A report that examines that the psychological, legal and technological strategy employed by leading organizations to reduce insider threats, a class of enterprise risk sustain by trusted person who has access to intellectual data, but uses that information outside of acceptable business requirements [15].

Malware Injection

Malware injection is a code, insert into cloud services that act as "valid instances" and run as SaaS to cloud servers. In cloud, Malware injection attacks an attacker to inject mischievous services or virtual machine into the cloud. To prevent cloud from malware injection attack we can merge the integrity with hardware. We can use hardware for integrity purpose because for an attacker it is difficult to interfere in the IaaS level [16].

Abuse of Cloud Services

The development of cloud-based services has made it possible for both small and enterprise-level organizations to host vast amounts of data easily. However, the cloud's outstanding storage capacity has also allowed both hackers and authorized users to easily host and spread malware, illegal software, and other digital properties. This practice affects both the cloud service provider and its customer [17].

For example: privileged users can directly or indirectly increase the security risks and as a result infringe upon the terms of use provided by the service provider. These risks include the sharing of pirated software, videos, music, or books, and can result in legal consequences in the forms of fines and settlements with the U.S. Copyright Law reaching up to $250,000. Depending on the damage, these fines can be even more cost prohibitive. You can reduce your exposure to risk by monitoring usage and setting guidelines for what your employees host in the cloud. Service providers and legal entities, such as CSA have defined what is abusive or inappropriate behavior along with methods of detecting such behaviors [17].

Insecure APIs

Application Programming Interfaces (API) gives user the opportunity to customize their cloud experience. But, APIs can be a threat to cloud security because of their very nature. Not only they give companies the ability to customize features of their cloud services to fit business needs, but they also authenticate, provide access, and effect encryption. APIs give programmers the tools to build their programs to integrate their applications with other job-critical software. Simple example of an API is YouTube, where developers have the ability to integrate YouTube videos into their sites or applications. The vulnerability of an API lies in the communication that takes place between applications. While this can help programmers and businesses, they also leave exploitable security risks [18].

Denial of Service Attacks

Denial of service (DOS attack) is cyber-attacks in which the offender seeks to make a machine or network source unavailable to its intended users by temporarily disrupting services of a host connected to the internet. DOS is typically accomplished by flooding the targeted machine or resources with superfluous requests in an attempt to overload systems and prevent some legitimate requests from being fulfilled. In some cases, however, DOS is also used as a smokescreen for other malicious activities, and to take down security appliances such as web application firewalls [11].

Insufficient Due Diligence

Most of the issues we've looked at here are technical in nature, but this particular security gap occurs when an organization does not have a clear plan for its goals, resources, and policies for the cloud. Insufficient due diligence- with cloud computing being a new implementation, especially to the hiring organizations, there is a knowledge gap that can prevent sufficient exercise of due diligence when hiring a

IIOAB JOURNAL
ISSN: 0976-3104

cloud service provider. In other words, it's the people factor. This is especially important to companies whose data falls under regulatory laws like PII, PCI, PHI, and FERPA or those that handle financial data for customers [13].

Shared Vulnerabilities

Cloud security is a shared responsibility between the service provider and the customer. This partnership between customer and provider requires the customer to take preventative actions to protect their data. While major providers like Box, Dropbox, Microsoft, and Google do have mass procedures to secure their side. As per article "Office 365 Security & Share Responsibility," the protection of user passwords, access restrictions to both files and devices, and multi-factor authentication – firmly in user hands. Hence the customers and service providers have shared responsibilities and omitting yours can result in your data being compromised [18].

Data Loss

Another serious threat is that an important data compromised due to deletion, modification, unlinking a record and storing of data on unreliable medium. It leads to loss of crucial data, reputation (for businesses), and trust of customers. Loss of data may cause severe legal and policy compliance issues. Malicious attack, natural disaster, or a data wipe by the service provider leads to a loss of data on cloud services. Losing important information can be devastating to businesses that don't have a recovery plan. In 2011 Amazon is an example of an organization that suffered data loss by permanently destroying many of its own customers' data [19].

TECHNIQUES TO SECURE DATA IN CLOUD

Authentication and Identity

Authentication of users and even of communicating systems is performed by various methods, but the most common is cryptography. Authentication of users takes place in various ways like in the form of passwords that is known individually, in the form of a security token, or in the form a measurable quantity like fingerprint. One problem with using traditional identity approaches in a cloud environment is faced when the enterprise uses multiple cloud service providers (CSPs). In such a use case, synchronizing identity information with the enterprise is not scalable. Other problems arise with traditional identity approaches when migrating infrastructure toward a cloud-based solution [20].

Data Encryption

If you are planning to store sensitive information on a large data store then you need to use data encryption techniques. Having passwords and firewalls is good, but people can bypass them to access your data. When data is encrypted it is in a form that cannot be read without an encryption key. The data is totally useless to the intruder. It is a technique of translation of data into secret code. If you want to read the encrypted data, you should have the secret key or password that is also called encryption key [20].

Information integrity and Privacy

Cloud computing provides information and resources to valid users. Resources can be accessed through web browsers and can also be accessed by malicious attackers. A convenient solution to the problem of information integrity is to provide mutual trust between service provider and user. Another solution can be providing proper authentication, authorization and accounting controls so that the process of accessing information should go through various multi levels of checking to ensure authorized use of resources. Some secured access mechanisms should be provided like RSA certificates, SSH based tunnels [20].

Availability of Information

Non-availability of information or data is a major issue regarding cloud computing services. Service Level agreement is used to provide the information about whether the network resources are available for users or not. It is a trust bond between consumer and service provider. A way to provide availability of resources is to have a backup plan for local resources as well as for most crucial information. This enables the user to have the information about the resources even after their unavailability [20].

Secure Information Management

It is a technique of information security for a collection of data into central repository. It is comprises of agents running on systems that are monitored and then sends information to a server that is called "Security Console". The security console is managed by admin who is a human being who reviews the information and takes actions in response to any alerts. As the cloud user base, dependency stack increase, the cloud security mechanisms to solve security issues also increase, this makes cloud security management much more complicated. It is also referred as a Log Management. Cloud providers also provide some security standards like PCI DSS, SAS 70. Information Security Management Maturity is another model of Information Security Management System [20].

Malware-injection attack solution

This solution creates a no. of client virtual machines and stores all of them in a central storage. It utilizes FAT (File Allocation Table) consisting of virtual operating systems. The application that is run by a client

can be found in FAT table. All the instances are managed and scheduled by Hypervisor. IDT (Interrupt Descriptor Table) is used for integrity checking [20].

CONCLUSIONS

Cloud computing is a promising and emerging technology for the next generation of IT applications. The barrier and hurdles toward the rapid growth of cloud computing are data security and privacy issues. Reducing data storage and processing cost is a mandatory requirement of any organization, while analysis of data and information is always the most important tasks in all the organizations for decision making. So no organizations will transfer their data or information to the cloud until the trust is built between the cloud service providers and consumers. A number of techniques have been proposed by researchers for data protection and to attain highest level of data security in the cloud. However, there are still many gaps to be filled by making these techniques more effective. More work is required in the area of cloud computing to make it acceptable by the cloud service consumers. This paper surveyed different techniques about data security and privacy, focusing on the data storage and use in the cloud, for data protection in the cloud computing environments to build trust between cloud service providers and consumers.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

This is to acknowledge to Dr. Pasenjit Bannerjee, Accendere Knowledge Management Services for his endless support for the successful completion of this paper. However, for any mistake the authors are solely responsible for that.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Behl A. [2011] Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation. 2011 World Congress on Information and Communication Technologies. IEEE, doi: 10.1109/WICT.2011.6141247.

[2] Behl A, Behl K. [2012] An analysis of cloud computing security issues. 2012 world congress on information and communication technologies. IEEE, doi: 10.1109/WICT.2012.6409059.

[3] Ertaul L, Singhal S, Saldamli G. [2010] Security Challenges in Cloud Computing. In Security and Management, doi: 10.1.1.722.4218.

[4] Mell P, Grance T. [2011]. The NIST definition of cloud computing. https://csrc.nist.gov/publications/detail/sp/800-145/final

[5] Mell P, Grance T. [2009] The NIST definition of cloud computing, version 15. National Institute of Standards and Technology (NIST). Information Technology Laboratory.

[6] Catteddu, D. [2009]. Cloud Computing: benefits, risks and recommendations for information security. In Iberic Web Application Security Conferenc). Springer, 10.1007/978-3-642-16120-9_9.

[7] Bhadauria R, Sanyal S. [2012] Survey on security issues in cloud computing and associated mitigation techniques. arXiv preprint arXiv:1204.0764.

[8] Jadeja Y, Modi K. [2012] cloud computing- concepts, architecture and challenges, IEEE, doi: 10.1109/ICCEET.2012.6203873.

[9] Balasubramanian R, Aramuthan DM. [2012]. Security problems and possible security approaches in cloud computing. Int J Sci Eng Res, 3(6), 1-4.

[10] Ukil A, Jana D, De S. [2013]. A security framework in cloud computing infrastructure. IJCSITS, 5(5), 11.

[11] Padhy RP, Patra MR, Satapathy SC. [2011]. Cloud computing: security issues and research challenges. IJCSITS, 1(2), 136-146.

[12] Dubey K, Kumar M, Chandra MA. [2015]. A priority based job scheduling algorithm using IBA and EASY algorithm for cloud metaschedular. In 2015 International Conference on Advances in Computer Engineering and Applications. IEEE, doi: 10.1109/ICACEA.2015.7164647.

[13] Srinivasan S, Raja K, Muthuselvan S. [2012]. Futuristic assimilation of cloud computing platforms and its services. In 2012 International Conference on Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM). IEEE, doi: 10.1109/ICETEEEM.2012.6494467.

[14] Bakshi A, Dujodwala YB. [2010, February]. Securing cloud from DDOS attacks using intrusion detection system in virtual machine. In 2010 Second International Conference on Communication Software and Networks. IEEE, doi: 10.1109/ICCSN.2010.56.

[15] Bakshi A, Dujodwala YB. [2010, February]. Securing cloud from DDOS attacks using intrusion detection system in virtual machine. In 2010 Second International Conference on Communication Software and Networks. IEEE, doi: 10.1109/ICCSN.2010.56.

[16] Williamson A. [2011] Comparing cloud computing providers. Cloud Comp J, 2(3):3-5.

[17] Zhang X, Wuwong N, Li H, Zhang X. [2010] Information security risk management framework for the cloud computing environments. In 2010 10th IEEE international conference on computer and information technology. IEEE, doi:10.1109/CIT.2010.501.

[18] Reddy VK, Rao BT, Reddy LSS. [2011] Research issues in cloud computing. Global Journal of Computer Science and Technology, 11(11):59-64.

[19] Lim HC, et al. [2009] Automated control in cloud computing: challenges and opportunities. In Proceedings of the 1st workshop on Automated control for datacenters and clouds. doi:10.1145/1555271.1555275.

[20] Habib SM, Ries S, Muhlhauser M. [2010] Cloud computing landscape and research challenges regarding trust and reputation. In 2010 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic and Trusted Computing, IEEE, doi: 10.1109/UIC-ATC.2010.48.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

TOWARDS THE STUDY OF AN ARTIFICIAL INTELLIGENCE AND ITS RESEARCH FIELDS

Varsha Deora[1], Srishty Jindal[1], Prateek Jain[2]*

[1]*Faculty of Engineering & Technology, Department of Computer Science, Manav Rachna International Institute of Research & Studies, Faridabad, INDIA*

[2]*Accendere CL Educate Ltd., New Delhi, INDIA*

ABSTRACT

AI being a revolutionary world has totally encapsulated our day to day lives. It makes the unique combination of minds & the machines. From the last couple of years there is gradual increase in the AI spreading its roots in all the fields such as machine learning, robotics, neural networks etc. Artificial intelligence is the intelligence exhibited by the computers. The word artificial means not natural or man-made and the word intelligence means the ability to acquire and apply knowledge and skills. There are many research fields of AI each having its own set of applications techniques and importance. This paper aims at focusing about the artificial intelligence at a glance and its applications/research fields. It helps in providing us the automation and hence leading to a very bright scope and future ahead.

KEY WORDS
Artificial intelligence, machine learning, computers, robots, machines.

Received: 26 Mar 2019
Accepted: 18 May 2019
Published: 2 June 2019

INTRODUCTION

Artificial intelligence is defined as the ability of computers to think and respond like humans and to perform complex tasks. Just like Wright Brothers imitated a bird's flight to build the architecture of first successful flying plane; artificial intelligence is an approach to imitate human intelligence. John McCarthy coined the term Artificial Intelligence in 1956 as a branch of computer science concerned with making of computers which behaved like humans.AI is an immerging field of research area due to its very usefulness in daily life and expert systems. AI mimics the human intelligence and improves the quality and speed of work without any human errors.

Artificial intelligence not only mimics human intelligence but also the other intelligent species in nature. One example of such intelligence is swarm intelligence. In this the nature of ant colony is studied where they don't have independent functioning but altogether in a group, they prove to be highly efficient system. AI is capable enough to perform tasks with high precision rates which a human cannot do. This man-made computer-based intelligence is showing us better results than human intelligence in general. It's is a combination of physiological reasoning and perception along with mathematical computations.AI can be broadly divided into two categories naming strong AI and weak AI. [1]. The types of AI is being shown in Fig 1.

Strong AI- A computer system is said to be a strong AI if it is at-least as smart as humans i.e. replicate human level of intelligence. It is capable of performing all tasks independently without any human interaction. It has capabilities to think, reason, perceive and judge. There are ethics for creating a strong AI. But it is believed that a strong AI is not yet created and if created can replace humans completely.

Weak AI- Self intelligent machines used to perform particular tasks, witness Expert Systems, robots, predictions and recommendations etc.

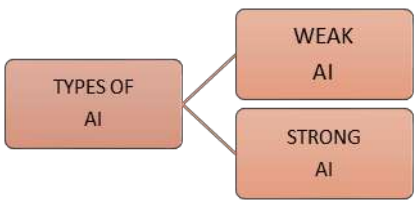

Fig. 1: Types of AI.

LITERATURE REVIEW

L.liu et al [2] projected a modified version of HMM called Self Adaptive HMM. A huge data set of 20,000 Chinese tweets were collected and classified on the basis of anger, happiness, sadness and fear. It uses Swarm Optimization algorithm to optimize the parameters used. The model decided whether the unlabeled tweet will be classified to which category after being assigned an emotion. The data set which was analyzed scored a 76% on happiness and fear and 65% on anger, surprise & sadness.

***Corresponding Author**
Email:
Prateek.jain@accendere.co.in

IIOAB JOURNAL
ISSN: 0976-3104

His research on self-adaptive HMM was proved to be better than Naïve Bayes & Support Vector Machine on identification of happiness, sadness &anger.

B.Yang et al [3] correctly identified type of land in terms of usage by remote sensing images. It combined several classifier systems like Support vector machine, maximum likelihood classifier, minimum distance classifier, back propagation neural networks and Fuzzy c-means to build a three multiple classifier system. They used the ground truth map to compare the results of 3 multiple classifier system and various other base classifier such as Bayesian average (BA). Performance of multiple classifier system was compared with overall accuracy of 94.2% which is higher than others.

S. R. Devi et al [4] various neural network models were compared to predict rainfall one-day advance of Nilgiri. The various neural network models used are cascade -forward back propagation neural network (CBPN), nonlinear autoregressive exogenous network (NARX), feed forward back propagation neural network (BPN) and distributed time delay neural network (DTDNN). Parameters used to measure are daily rainfall, humidity and temperature and their forecasting capabilities are compared. Gradient Descent Graphs are used to do so. The data is collected from 14 rain gauge stations located near Niligiri. According to performance analysis nonlinear autoregressive exogenous network (NARX) prove d to be the best forecasting model.

M. Sahami et al [5] email filtering was performed to separate spam mails from important messages. The important mails and spam were classified accordingly. Initially this might seem to be a straight forward text classification problem, but they showed that by including domain specific features of this problem along with text and words, they can improve the classifier and produce accurate results of filtration. Therefore they displayed this problem as a decision theoretic framework and used probabilistic learning methods along with differential misclassification cost to produce better filters. They also showed the efficiency of such filters in practical and real life applications.

Turing test [6]

Turing test is used to test how intelligent an AI is. The test is conducted between two humans and a machine. The one human acts as a judge and test the other two on certain parameters by asking questions in text format. If the judge is unable to differentiate between the human and the machine, it is concluded that the machine is a good AI and has the ability to think just like humans can.

CAPTCHA- “Completely Automated Public Turing Test to tell Computers and Humans Apart” is a test developed by Alan Turing to distinguish between human users and robots. It includes distorted text and numbers which can easily be read by humans but difficult for robots to read and prevents robots from access.

GOALS OF AI

Artificial intelligence central goal is to make computers more useful and improve lifestyle. It is built for easy execution of complex tasks. The long-term goal is to study the existing intelligent systems and understand human working techniques and improving it. Providing more accurate results to the problems stated through logic and reasoning. To create correct results AI needs to have extensive knowledge of the world and represents various relationships, objects. In order to achieve the goals, the agents require a Path to be followed which is well planned and defined. To be able to make predictions one need to visualize the future i.e. has a perception of what the expected result is. And to obtain these results the machine needs to be trained which is also known as machine learning. The various goals of an AI are depicted in Fig 2 below.

Fig. 2: Goals of AI.

IIOAB JOURNAL
ISSN: 0976-3104

Many machines these days use not only software interface but also hardware interface and responds in the real time environment. Such machines which interact with the real time environment are called robots. And we need motion and manipulations of these robots.

ROOTS OF ARTIFICIAL INTELLIGENCE

As Artificial intelligence is based on the various disciplines such as computer science, biology, psychology, physiology, math's, computation and engineering. All of this basis doesn't always work independently but are overlapped to create a good AI. Fig 3 denotes roots of an AI.

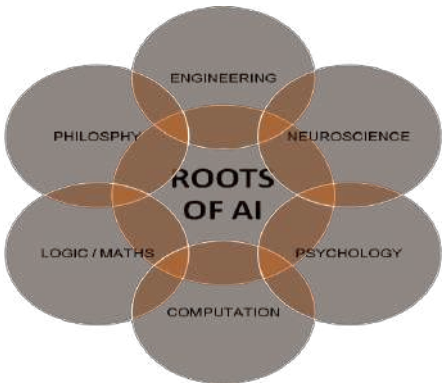

Fig. 3: Roots of AI.

Advantages and disadvantages of AI

Artificially intelligent systems mimic the human intelligence but unlike humans the results are not affected by emotions and there are no distractions to perform a task. A computer system can perform task at much faster speed than a typical human being and is also faster than the traditional methods used and unlike humans a machine doesn't gets tired and has the ability to work continuously without any breaks. AI requires large training set and knowledge which in case of humans will take forever to learn but in case of AI the once a machine is trained the knowledge can be transferred quickly.

Apart from the advantages over human beings artificial intelligent systems also perform better than traditional computing methods because traditional computing methods work on specific problems whereas the AI works on generic problems.

With the rise of artificial intelligence and increasing research it also has some disadvantages. Cost of development of an AI is huge and the skills require to build a good AI is vast. Due to the increasing use of robots in our daily life it is replacing human work and thus creating unemployment. On the other hand it is believed that computer can never replace human interaction and high posting jobs like care given by nursing in medical care, police enforcement and judges. There is a lack of common sense. It is believed that if we are successful in building a strong AI which can completely replace humans it can overpower humans and if in wrong hands can lead to mass scale destruction of humanity.

FIELDS OF ARTIFICIAL INTELLIGENCE

There are several research fields in AI each with a different purpose and set of applications.

Machine learning [7] is one such field of artificial intelligence that mimics the human abilities by machine. Both explanation and experienced based learning are parts of ML. machine these days are replacing human labour. But the question arises of how to train these machines according to our requirements. Instead of building heavy machinery with explicit programming, different algorithms are introduced which will help the machine to understand the virtual environment and take decisions. Because of this machines have become independent and have capabilities to take decisions on their own. Machine learning can be classified into 3 broad categories on the basis of nature of learning. These are:

Supervised learning- given the "right answer" for each example. The computer is presented with example training sets and their desired output is mapped, just like a "teacher" super vises the student to do work. The goal is to learn a general rule that maps inputs to outputs.

Unsupervised learning-No labels associated with it. It is algorithm's duty to find some structure in the data set for us. Unsupervised learning can itself be a goal in discovering hidden patterns in the data. Clustering is an example of unsupervised learning.

Reinforcement learning- it is inspired by behavioral psychology. A computer program is made to interact with environment in which it must perform a certain goal and on achieving the goal it is awarded for the good response.

Some of the machines learning algorithms are linear regression, decision tree learning, association rule learning, similarity and matrix learning.

Applications for machine learning include: Bioinformatics, Classifying DNA sequences, Computational anatomy, Computer vision, including object recognition , Detecting credit card fraud, Game playing, Information retrieval, Internet fraud detection [8], Marketing, Machine learning control, Medical diagnosis, Economics, Natural language understanding, Online advertising, Recommender systems[9], Robot locomotion, Search engines, Sentiment analysis, Software engineering, Speech and handwriting recognition, Financial market analysis, Structural health monitoring.

Neural networks are another field of artificial intelligence. An artificial neural network (ANN) [10] is based upon the structural and functional aspects of the biological neural networks. Computations are structured in terms of interconnected groups of artificial neuron mimicking the functioning of neurons in human body, processing information. Modern neural networks are non-linear statistical data modelling tools. There are many layers to it called as nodes and like a neuron transmits the da ta to our brain nodes work in a similar fashion.

Deep learning consists of multiple hidden layers of artificial neural networks (Fig-4).

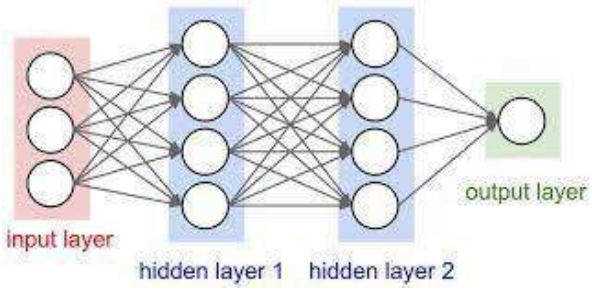

Fig. 4: Deep learning.

Logic-Fuzzy logic is a method of reasoning which resembles human reasoning. In digital world we have only two discrete possibilities i.e. 0/1 or true/false or yes/no. but in real world there are many more possibilities like maybe, certainly, possibly, greatly, surely etc. there are several application of fuzzy systems in our daily life including automotive systems–aviation, automatic gearboxes, four-wheel steering vehicle environment, environment control- air conditioners /dryers /heaters, humidifiers, microwave ovens, refrigerators, toasters, vacuum cleaners, washing machines.

Expert System [11] In artificial intelligence, an expert system is a computer system that copies the decision-making ability of a human professional. Expert systems are basically used for solving the complex problems through reasoning about the knowledge. An expert system can be classified in two parts: the inference engine and the knowledge base. The knowledge base represents rules & facts. The inference engine applies the rules to the known facts to infer new facts. Inference engines can also include justification and debugging facility. Apart from this expert system are capable of Recommending, directing and supporting human in decision making, Representing, obtaining a solution, Diagnosing, predicting results, Suggesting alternative options to a problem.

Applications of expert systems are Diagnosis Systems to realize cause of disease from experiential data, conduction medical operations on humans. Evaluating data continuously with observed system or with approved behavior such as leakage monitoring in long petroleum pipeline. Recognition of possible fraud, doubtful transactions. Robotics [Fig-5]–robots are the artificial agents acting in real world environment and reducing man power from performing certain tasks. The other AI fields work mostly on software whereas the robotics requires separate hardware. There is movement in a robot which helps it to interact with the real world. The higher the number of movements in a robot the greater will with the number of possible movements and much more complexity. Robots are mainly used in industries to perform heavy tasks for which large human force will be required. Along with industries robots are also used in medical treatments to perform operations requiring high precision. Other than this robot are used for many scientific purposes like mars rover. It can also be used just for entertainment purposes like camera drones.

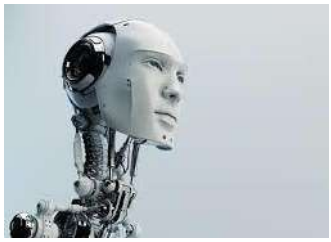

Fig. 5: Robotics [12].

Natural language processing it is a field of computational linguistics and artificial intelligence. The main focus of NLP is to understand the natural languages spoken by humans like English and grammar by the machines. This understanding provides a medium of communication between the intelligent systems and the humans. A set of ML algorithms and models can be used to achieve NLP based AI.

Some of the aspects of NLP are translation, classification, clustering, and information extraction [Fig-6].

Fig. 6: Steps involved in natural language processing.

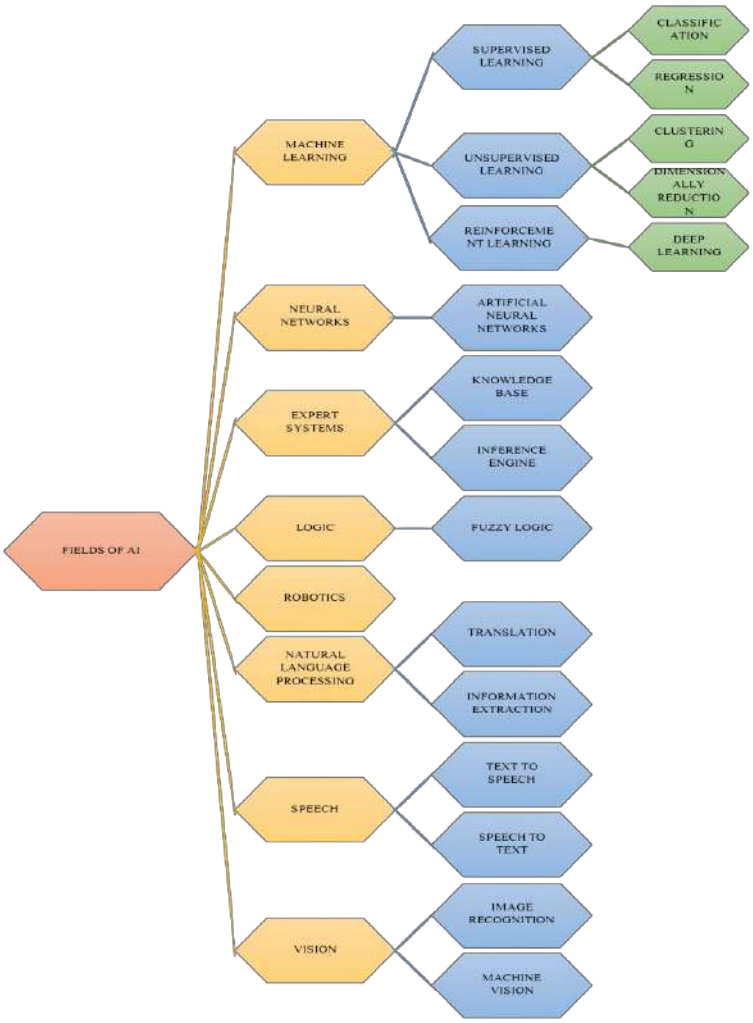

Fig. 7: Fields of AI.

Speech

Speech recognition is a important field these days. It includes speech to text conversion and text to speech conversion. Speech recognition is different from voice recognition as speech recognition is user independent and only focuses on the words being spoken and not the person speaking [13].

Vision

Image recognition and machine vision are two major aspects of vision. Remote sensing using images is done by this method. Snapchat filters also use face detection which is a part of image recognition. In the fields of robotics machine vision is used along with sensors to guide the path to the robot for [14,15] movement. The fields of AI is represented in Fig-7.

CONCLUSION

From the last couple of decades, a drastic rise is being seen in every area of the AI including robots, automatic vehicles etc. It has completely transformed the world by making every aspect so easy especially by the help of the machine learning principles/techniques. The current progress in AI is just the beginning of the future trend. Now ('vicarious') AI is so advanced that it can surpass the CAPTCHA test. Latest advancements in robotics include human like robot SOPHIA made by Hanson robotics capable of having full conversations, showing empathy, understanding emotions and responding to them and also having facial expressions. In the hypothetical future there will be singularity where the super -intelligent machines far beyond the capability of humans will exist. Morality could determine the ethics of AI. Being advancement, it is giving an automated path to us leading to a very bright future & things like drones, humanoid robot that can ride a motorcycle & self-driving car are becoming the main driver of innovation for an automotive industry. Along with it a lot of employment opportunities also exists with this technique.

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None.

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] Borana J. [2015] Applications of Artificial Intelligence & Associated Technologies. Proceeding of International Conference on Emerging Technologies in Engineering, Biomedical, Management and Science [ETEBMS-2016], 64-67.

[2] Liu L, et al. [2014] A Self-Adaptive Hidden Markov Model for Emotion Classification in Chinese Microblogs, Hindawi Publishing Corporation Mathematical Problems in Engineering. 2015, Article ID 987189.

[3] Yang L, et al. [2015] Automatic Classification of Remote Sensing Images Using Multiple Classifier Systems, China.

[4] Devi S, et all. [2016] Performance Comparison of Artificial Neural Network Models for Daily Rainfall Predictio, International Journal of Automation and Computing.

[5] Saham M, et al. [1998] In: AAAI-98 Workshop on Learning for Text Categorization. Madison, Wisconsin.55–62, AAAI Technical Report WS-98-05.

[6] Pann A. [2015] Artificial Intelligence and its Application in Different Areas, International Journal of Engineering and Innovative Technology. 4(10): 79-84.

[7] Talwa A. [2013] Machine Learning: An artificial intelligence methodology, International Journal Of Engineering And Computer Science, 2(12):3400-3404.

[8] McClendon L, Meghanatha N. [2015] Using machine learning algorithms to analyze crime data. Machine Learning and Applications: An International Journal, 2(1). doi: 10.5121/mlaij.2015.2101.

[9] Bhat B. [2014] A Review Paper on Machine Learning Based Recommendation System, IJEDR. 2(4): 3955-3961.

[10] Sawale GJ. [2013] Use of Artificial Neural Network in Data Mining For Weather Forecasting. International Journal Of Computer Science And Application. 6(2): 383-387.

[11] Khanna S. [2010] Expert Systems Advances in Education NCCI 2010 -National Conference on Computational Instrumentation CSIO Chandigarh, INDIA. 19-20.

[12] Chaudhary KR. [2012] Goals, Roots and Sub-fields of Artificial Intelligence. MBM Engineering College, Jodhpur, India.

[13] Xu K. [2011] Dynamic neuro-fuzzy control design for civil aviation aircraft in intelligent landing system. in: 2011 IEEE International Conference on Mechatronics and Automation, IEEE, doi: 10.1109/ICMA.2011.5986355

[14] Deepa SN, Aruna DB. [2011] A survey on artificial intelligence approaches for medical image classification. Indian Journal of Science and Technology. 4(11): 1583-1595.

[15] Vaidya OS, Kumar S. [2006] Analytichierarchy process: an overview of applications, European Journal of Operational Research. 169(1):1-29.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

HAND AND GESTURE RECOGNITION TECHNIQUES

Himanshu Miglani*, Hari Mohan Sharma, Agha Imran Husain
Department of CSE, Manav Rachna Univeristy, Faridabad, INDIA

ABSTRACT

Gestures help to communicate in a natural way between user and computer system in a virtual environment. Hand gesture is a method of non-verbal communication for human beings to express much more other than their body parts. Hand gesture recognition has a great role in designing an efficient human computer interaction system. In this report, a survey of various gesture recognition approaches is provided with particular emphasis on hand gestures.

KEY WORDS
HCI,HMM,FSM,GAs,ANN, Oriented Histogram,Cluster Algorithms

INTRODUCTION

Gestures and facial expressions can be used to communicate with the computers which require the computer system to understand and analyze the signals to perform a particular movement [2]. Recently the designing of special input devices proved to facilitate the interaction between humans and computers. Gesture recognition has been applied in a large range of application areas such as recognizing sign language, human computer interaction (HCI), robot control, smart surveillance, lie detection, visual environments manipulating, etc. Now a days different techniques and tools have been used for handling gesture recognition that vary between mathematical models like Hidden Markov Model (HMM) and Finite State Machine (FSM) to approaches based on software computing methods such as fuzzy clustering, Genetic Algorithms (GAs) and Artificial Neural Network (ANN). Since human hand is a complex articulated object which is controlled by 35 muscles and requires 27 degrees of freedom to be versatile in all the movements, it is a thrust area of research [3]. In today' digital field implementing gesture recognition system requires different type of devices such as cameras, instrumented gloves and coloured markers.

HAND GESTURE TECHNOLOGY

For any technology to come in action it first need to collect data to accomplish a specific task. For hand gesture recognition system different technologies are used for acquiring data. Present technologies that are used for recognizing gestures can be divided into vision based, instrumented glove and coloured marker approaches.

Received: 20 Mar 2019
Accepted: 11 May 2019
Published: 4 June 2019

Vision based approach

In this type of approach the system requires only camera to capture images required for the interaction between human and computers and no extra device is needed as shown in Fig-1. Although this approach is simple but a lot of challenges rose such as the complex background, lightning variation and other skin colour objects with the hand objects. Beside this system requires various measurements such as velocity, recognition time, robustness and computational efficiency [4][5].

Fig.1: Vision based Hand Gesture.

Instrumented glove approach

This type of glove uses sensor devices for capturing hand postures, motion and position. This approach can provide accurate coordinates of palm and finger's location and orientation and hand configurations. However the approach requires the user to connect with the computer physically which is a barrier

***Corresponding Author**
Email:
hmiglani78@gmail.com

between the interaction of computer and the user as shown in Fig- 2. The price of this instrumented glove is quite high and is inefficient for working in virtual reality [5] [6] [7].

Fig.2: Glove based approach.

Colored marker approach

In this type of approach marked gloves or colored markers are gloves that are worn by human hands with some colors to direct the process of tracking the hand and location of the palm and fingers which provide the ability to extract the geometric features necessary to form the hand shape [8]. The color glove shape might consist of small regions with different colors or as applied in where different colors are used to represent the fingers and palms, where a wool glove was used as shown in Fig-3.

Fig.3: Color Marker.

VISION BASED HAND GESTURE RECOGNITION APPROACHES

Vision based technologies use a hand to extract the information needed for recognition of hand, these methods are natural [8] [9], easy, and the user directly interacts with the system. Vision based technology deals with some image characteristics such as texture and color for acquiring data needed for gesture analysis. There are many techniques that are applied for detecting hand object after some image pre-processing operations [9]. These methods can be divided into two parts.

Appearance based approaches

In this type of approach, visual appearance of input hand image is modelled using the feature extracted from stored image. Appearance based approaches are simpler and easier than 3D model based approaches due to the easier extraction of features in a 2D image. The common method used in this approach is to detect the skin coloured regions in the image; however this method is affected by changing illumination conditions and other background objects with skin like color.

3D model based approaches

This type of approach used 3D model description for modelling and analyzing the hand shape. In these approaches search for the kinematic parameters are required by making a 2D projection from 3D model of the hand to correspond edges images of the hand but a lot of hand features might be lost in 2D projection. 3D Model can also be classified into volumetric and skeletal models. Volumetric models deal with 3D visual appearance of human hand and usually used in real time applications. The main problem with this modelling technique is that it deals with all the parameters of the hand which are huge dimensionality. Skeletal models overcome the volumetric hand parameters problem by limiting the set of parameters to model the hand shape from 3D structure as shown in Fig-4.

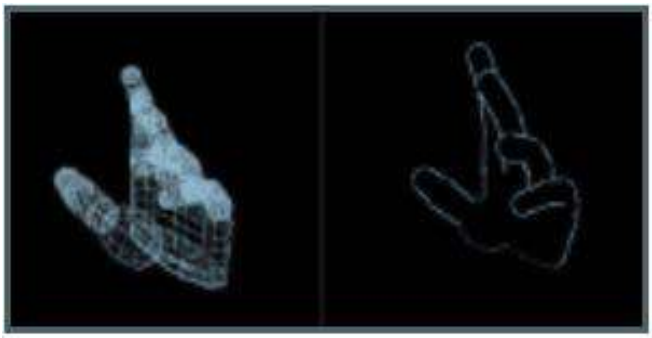

Fig. 4: 3D Approach.

GESTURE RECOGNITION TECHNIQUES

In this, the recognition of gesture involves several concepts such as pattern recognition, motion detection & analysis and machine learning [10]. For these different tools and techniques are utilized in a gesture recognition systems such as computer vision, image processing, pattern recognition and statistical modelling [9].

Artificial neural networks (ANN)

The use of neural network for gesture recognition has been examined by many researchers. Most of the researches use ANN to classify gesture recognition process, while some others use it to extract the shape of the hand. Maung (2009) presented a system for hand tracking and gesture recognition using Neural Network System to recognize Myanmar Alphabet Language (MAL). Adobe Photoshop filter is then used to find the edges of the input image and the histogram of local orientation employed to extract the image feature vector which would be an input to the supervised neural network system. Maraqa & Abu-Zaiter (2008) used two recurrent neural network architectures to recognize the Arabic Sign Language (ArSL). Elman partially recurrent neural network and fully recurrent neural networks have been used separately. A coloured glove used for input image data, and for segmentation process, the HSI colour model is applied. The segmentation divides the image into six colour layers, one for the wrist and five for fingertips. 30 features are extracted and grouped together to represent a single image. Fifteen elements are used to represent the angles between the fingertips and between them and the wrist, and fifteen elements to represent distances between fingertips; and between the fingertips and the wrist. This input feature vector is the input to both neural network systems. 900 coloured images were used as training set, and 300 coloured images for system testing. Results had shown that fully recurrent neural network system with recognition rate 95.11%, which is better than the Elman neural network with a 89.67% recognition rate.

Histogram based feature

Many researches have been applied based on the histogram, where the orientation histogram is used as a feature vector.

The first implementation of the orientation histogram in the field of the gesture recognition system and real time was performed by Freeman & Roth (1995), they presented a method for recognizing a gesture based on pattern recognition using orientation histogram. For digitized input image, black and white input video was used, some transformations were made in the image to compute the histogram of the local orientation of each image, then a filter is applied to blur the histogram, and plot it in polar coordinates. The system consists of two phases training phase and running phase. In the training phase, for different input gestures the training set is stored with their histograms. In running phase input image is presented to the computer and then the feature vector for the new image is formed, Then comparison performed between the feature vector of the input image with the feature vector (oriented histogram) of all images of the training phase using Euclidean distance metric and the less error between the two compared histograms will be selected. The total processing time was 100 millisec per frame.

Fuzzy clustering algorithm

Clustering Algorithm is a simple term that comprises of all methods that divide the given set of sample data into subsets or clusters. According to this the pattern that share same characteristics are grouped together to form a cluster. Xingyan Li (2003) Presented fuzzy clustering algorithm to recognize hand gestures in a mobile remote. A camera was used to acquire input raw images. The input RGB images are converted into an HSV colour model, and then the hand is extracted after some pre-processing operations to remove noise and unwanted objects and then thresholding using to segment the hand shape. 13

elements were used as a feature vector, first one for aspect ratio of the hand's bounding box, and the rest 12 parameters represent grid cell of the image, and each cell represents the mean gray level in the 3 by 4 blocks, partition of the image, where the mean value of each cell represents the average brightness of those pixels in the image.

Hidden markov model (HMM)

Many researches were applied in the field of gesture recognition using HMM. HMM is a stochastic process with a finite number of states of the Markov chain and a number of random functions so that each state has a random function. Keskin, Erkan & Akarun (2003) presented HCI interface based on real time hand tracking and 3D gesture recognition using hidden Markov models (HMM). Two coloured cameras for 3D construction are used. To overcome the problem of using skin colour for hand detection because of hand overlapping with other body parts markers is used to reduce the complexity in hand detection process. Markers used to segment the hand from complex backgrounds under invariant lighting conditions. The markers are distinguished using marker detection utility and a connected components algorithm was applied to find marker regions using double thresholding. For fingertip detection, simple descriptors were used where the bounding box and four outmost points of the hand that defining the box is determined. The bounding box in some cases needs to be elongated to determine the mode of the hand, and the points used to predict the fingertip location in different modes of the hand. Kalman filter was used for filtering trajectory of the hand motion. For 3D reconstruction of finger coordinates, calibration utility was implemented for the specific calibration object.

IMPLEMENTATION TOOLS

A lot of implementation, hardware and software tools have been used for recognizing gestures depending on the application fields they are used.

Hardware implementation tools

The input devices used in a gesture recognition system are many and different according to system and application used in the recognition process. In this only a single camera can be used for posture recognition since this environment might be inconvenient for other types of image based recognition. Stereo cameras that consist of two lenses with an isolated sensor for each lens which imitates human visual system, therefore, the 3D effect of views is created. Stereo cameras can be used to make 3D pictures for movies or for range imagining [28]. In this tracking device such as instrumented data gloves measure the finger movements through many types of sensors. It provides accurate information about the position and orientation of the hands using magnetic or inertial tracking devices [6][7]. In Controller-based gestures, controllers represent a complement of the human so that when body moves and creates some gestures, these motions are captured using some software. Mouse gesture is an example of such controllers [28].

Softwares implementation tools

Software tools are programming language and windows system that are used for implementing the gesture recognition system. Some research has used programming languages like C, C++ and Java language. Then to simplify the work, especially when image processing operations are needed, MATLAB ® with image processing toolbox is used.

APPLICATION AREAS OF HAND GESTURES SYSTEM

The hand gestures recognition system has been applied to different applications on different areas, as it is mentioned in [7][9] including; sign language translation, virtual environments, smart surveillance, robot control, medical systems etc. Some of the hand gesture application areas are mentioned below-

Sign language recognition

Since the sign language is used for interpreting and explanations of a certain subject during the conversation, it has received special attention [9]. A lot of systems have been proposed to recognize gestures using different types of sign languages [17]. For example, recognizing American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming matching. [20] recognized Japanese sign language JSL using Recurrent Neural Network, 42 alphabets and 10 words. [19] recognized Arabic Sign language ArSL using two different types of Neural Network, Partially and Fully Recurrent neural Network.

Graphic editor control

The graphic editor control system requires the hand gesture to be tracked and located as a preprocessing operation [23][24] used 12 dynamic gestures for drawing and editing graphic system. Shapes for drawing are; triangle, rectangular, circle, arc, horizontal and vertical line for drawing, and commands for editing graphic system are; copy, delete, move, swap, undo, and close.

Virtual environments (VEs)

One of the popular applications in gesture recognition system is virtual environments VEs, especially for communication media systems [7]. [25] provided 3D pointing gesture recognition for natural human computer Interaction HCI in a real-time from binocular views. The proposed system is accurate and independent of user characteristics and environmental changes.

Number recognition

Another recent application of hand gesture is recognizing numbers. [26] proposed an automatic system that could isolate and recognize a meaningful gesture from hand motion of Arabic numbers from 0 to 9 in a real time system using HMM.

Television control

The hand postures and gestures are used for controlling the Television device [18]. In [27] a set of hand gesture is used to control the TV activities, such as turning the TV on and off, increasing and decreasing the volume, muting the sound, and changing the channel using open and close hand.

3D modeling

To build a 3D model, a determination of hand shapes are needed to create, built and view the 3D shape of the hand gesture[18]. Some systems build the 2D and 3D objects using hand outlines. 3D hand modeling can be used for this purpose also which still a promising field of research.

CONCLUSION

Building an efficient human machine interaction is an important goal of the gesture recognition system. Many applications of a gesture recognition system range of virtual reality to sign language recognition and robot control. In this research, a survey of various tools and techniques of gesture recognition system has been provided with an emphasis on hand gesture expressions. The major tools surveyed include HMMs, ANN and fuzzy clustering have been reviewed and analyzed. Most researchers use coloured images for achieving better results. Comparison between various gesture recognition systems have been provided with explaining the important parameters needed for any recognition system which include segmentation process, feature extraction and the classification algorithm.

Still, there are many conditions which are needed to be fulfilled. The first and the foremost thing which needs to be done are to enhance the recognition capability for various lighting conditions when using cameras. Then there is a need to obtain accuracy in the field of hand gesture recognition. There is need to implement a number of gestures and to identify them and applying gesture recognition for accessing internet applications.

CONFLICT OF INTEREST

None.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Accendere Knowledge Management Services for their valuable comments that led to substantial improvements on an earlier version of this manuscript.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Ibraheem NA, Khan RZ. [2012] Survey on various gesture recognition technologies and techniques. International journal of computer applications. 50(7).

[2] Daugman J. [1997] Face and Gesture Recognition: Overview, in: IEEE Transactions on Pattern Analysis and Machine Intelligence. 19(7): doi: 10.1109/34.598225

[3] Vladimir I, et al. [1997] Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. IEEE Transactions On Pattern Analysis And Machine Intelligence, 19(7):677- 695.

[4] Murthy GRS, Jadon RS. [2009] A Review if Vision Based Hand Gestures Recognition. International Journal of Information Technology and Knowledge Management, 2(2):405-410.

[5] Garg P, et al. [2009] Vision Based Hand Gesture Recognition. World Academy of Science, Engineering and Technology, 49:972-977.

[6] Dipietro L, et al. [2008] Survey of Glove-Based Systems and their applications, IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews. 38(4):461-482. doi: 10.1109/TSMCC.2008.923862]

[7] Joseph J, La Viola Jr. [1999] A Survey of Hand Posture and Gesture Recognition Techniques and Technology, Master Thesis, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA.

[8] Mokhtar M, et al. [2012] Hand Gesture Modeling and Recognition using Geometric Features: A Review, Canadian Journal on Image Processing and Computer Vision. 3(1): 1

[9] Mitra S, Acharya T. [2007] Gesture Recognition: A Survey. IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews. 37(3):311-324. doi: 10.1109/TSMCC.2007.893280

[10] Wu Y, et al. [1999] Vision-Based Gesture Recognition: A Review, Lecture Notes in Computer Science, Gesture Workshop, proceedings of the International Gesture Workshop on Gesture-Based communication in Human-Computer interaction. 1739:103-115.
[11] Hninn T, et al. [2009] Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks, World Academy of Science, Engineering and Technology. 50:466- 470.
[12] Maraqa M, et al. [2008] Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks, IEEE First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT. 478-48. doi: 10.1109/ICADIWT.2008.4664396
[13] William T, et al. [1995] Orientation Histograms for Hand Gesture Recognition, IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich.
[14] Li X. [2003] Gesture Recognition based on Fuzzy C-Means Clustering Algorithm, Department of Computer Science. The University of Tennessee. Knoxville.
[15] Keskin C, et al. [2003] Real Time Hand Tracking and 3D Gesture Recognition for Interactive Interfaces using HMM, In Proceedings of International Conference on Artificial Neural Networks.
[16] Mitra S, Acharya T. [2007] Gesture Recognition: A Survey IEEE Transactions on systems. Man and Cybernetics, Part C: Applications and reviews, 37(3):311-324. doi: 10.1109/TSMCC.2007.893280.
[17] Wysoski, et al. [2002] A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks. in: Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.doi: 10.1109/ICONIP.2002.1199054.
[18] Joseph J, LaViola Jr. [1999] A Survey of Hand Posture and Gesture Recognition Techniques and Technology, Brown University Providence, RI, USA
[19] Maraqa M, et al. [2008] Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks, IEEE First International Conference on the Applications of Digital Information and Web Technologies, (ICADIWT). 478-48. doi: 10.1109/ICADIWT.2008.4664396
[20] Kouichi M, Hitomi T. [1999] Gesture Recognition using Recurrent Neural Networks ACM conference on Human factors in computing systems: Reaching through technology (CHI '91). 237-242. doi: 10.1145/108844.108900
[21] Xingyan Li. [2003] Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm, Department of Computer Science. The University of Tennessee Knoxville
[22] Malima A, Ozgur E, Cetin M. [2006] A Fast Algorithm for Vision-Based Hand Gesture Recognition For Robot Control, IEEE 14th conference on Signal Processing and Communications Applications. 1-4. doi: 10.1109/SIU.2006.1659822
[23] Mitra S, Acharya T. [2007] Gesture Recognition: A Survey IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews. 37(3):311- 324. doi: 10.1109/TSMCC.2007.893280.
[24] Min B, et al. [1997] Hand Gesture Recognition Using Hidden Markov Models. IEEE International Conference on computational cybernetics and simulation. 5. Doi: 10.1109/ICSMC.1997.637364.
[25] Guan Y, Zheng M. [2008] Real-time 3D pointing gesture recognition for natural HCI. IEEE Proceedings of the 7th World Congress on Intelligent Control and Automation WCICA, doi: 10.1109/WCICA.2008.4593304
[26] Mahmoud E, et al. [2008] Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition. International Scholarly and Scientific Research & Innovation, 2(5):985-992.
[27] Freeman WT, Weissman CD. [1995] Television Control by Hand Gestures. IEEE International Workshop on Automatic Face and Gesture Recognition. Controller-based-gestures.
[28] https://en.wikipedia.org/wiki/Gesture_recognition

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

A REVIEW OF PULSE COUPLED NEURAL NETWORK

Nilansh Dewan*, Vaibhav Kashyap, Anup Singh Kushwaha
Department of CSE, Manav Rachna Universitry,Faridabad, INDIA

ABSTRACT

Multimodal therapeutic picture combination, as a viable gadget for the clinical applications, has made with the approach of various imaging modalities in therapeutic imaging. The essential motivation is to get most pertinent information from sources into a singular yield, which accept a basic part in remedial finding. In this paper, we examine about Pulse Coupled Neural Networks (PCNN) in Multimodal Image combination, The different patterns in PCNN and the employments of PCNN. Pulse-coupled neural systems (PCNN) in our condition. PCNN is a visual cortex-propelled neural system and portrayed by the worldwide coupling and heartbeat synchronization of neurons. It has been demonstrated reasonable for picture handling and effectively utilized in picture combination. In this Paper, we will discuss about the various applications of PCNN.

KEY WORDS
PulseCoupled Neural Network(PCNN), Magnetic Imaging Resonance(MRI),Computed Tomography(CT),

INTRODUCTION

Artificial Neural Networks in Change Detection ANNs give a channel for dealing with the intricate information viable. Like some other strategies ANN likewise has its related issues: A reasonable ANN engineering with a particular number of concealed hubs, required learning rate, energy, esteem, adds up to a number of an emphasis required, and the encoding strategy to signify the information and yield information ought to be resolved. Regularly a trail and blunder based procedure is received to discover these parameters. VHR pictures give a more prominent level of understanding into the transient land cover changes, meanwhile taking care of considerable quantum of data and overseeing the related blunders in enlistment and characterization is a test. In a solitary neural system engineering for VHR pictures with two particular stages to perform. Arrive cover change examination is proposed. The proposed neural design can, all the while misuse multi-phantom and multi-worldly data that are identified with the pixel ghostly reflectance changes and can create the last yield delineate blending three Neural Network comes about. [1]. The key preferred standpoint of this system is its capacity to identify the progressions and perceive the kind of class change. In another measurement to change recognition is given utilizing PCNNs. A particular mark of the scene is made from the swells produced by every cycle and is thinking about progressively to generate the change yield outline. The striking part of PCNNs is that it can use the logical and unearthly data at a similar occasion, which settles on it as a perfect decision for change recognition utilizing VHR pictures. Additionally, PCNNs don't need to go the data through numerous layers as done on account of ANNs. PCNNs have just a single layer of neurons that procedure the info got specifically from the first picture and to produce the subsequent heartbeat yield picture.

Received: 27 Mar 2019
Accepted: 19 May 2019
Published: 9 June 2019

As of date, therapeutic imaging has pulled in extending thought on account of its fundamental part in human administrations. Regardless, remarkable sorts of imaging methodology, for instance, X-shaft, enrolled tomography (CT), appealing resonation imaging (MRI), alluring resonation angiography (MRA), et cetera., give obliged information where a little information is ordinary, and some are novel. For example, X-pillar and registered tomography (CT) can give thick structures like bones and embeds with less contortion, however, it can't distinguish physiological changes. Thus, ordinary and neurotic delicate tissue can be better pictured by the MRI picture though PET can be utilized to give better data on the bloodstream and surge action with low spatial determination. Therefore, the anatomical and useful medicinal pictures should have been joined for a succinct view.

PCNN is also essential for Remote Sensing applications like Synthetic aperture radar (SAR) imaging plays an important role in both civilian life and military defense., SAR image fusion, which may become a fundamental procedure for the classification, detection, and recognition of targets in SAR images, has been attracting increased research interest worldwide. Synthetic-aperture radar (SAR) is a form of radar that is used to create two- or three-dimensional images of objects, such as landscapes. SAR uses the motion of the radar antenna over a target region to provide finer spatial resolution than conventional beam-scanning radars. SAR is typically mounted on a moving platform, such as an aircraft or spacecraft, and has its origins in an advanced form of side-looking airborne radar (SLAR). The distance the SAR device travels over a target in the time taken for the radar pulses to return to the antenna creates the large synthetic antenna aperture (the size of the antenna). Typically, the larger the aperture, the higher the image resolution will be, regardless of whether the aperture is physical (a large antenna) or synthetic (a moving antenna) – this allows SAR to create high-resolution images with comparatively small physical antennas. As a result of Earth observation satellites, such as ERS-1/2 (European Remote Sensing), ENVISAT (environment satellite), Radarsat 2 and Terra Sar X, a large number of SAR images are now available [2].

***Corresponding Author**
Email:
nilanshd@gmail.com

IIOAB JOURNAL
ISSN: 0976-3104

MULTIMODAL MEDICAL IMAGE FUSION

The investigation of multimodal restorative picture combination has pulled in more consideration because of the expanding requests of clinical applications. Restorative picture combination helps doctors remove the components that may not be ordinarily obvious in pictures by various modalities (e.g., MRI-T1 gives more noteworthy detail of anatomical structures, while MRI-T2 gives more noteworthy differentiation amongst ordinary and unusual tissues). Along these lines, keeping in mind the end goal to concentrate more data, medicinal picture combination consolidates these differentiating and complementary components into one melded picture. Therapeutic picture combination helps in diagnosing illnesses, as well as lessens the capacity taken a toll by decreasing stockpiling to a solitary melded picture rather than various source pictures. Individuals have proposed many methodologies, for example, FSD pyramid, inclination pyramid, Laplacian pyramid. DWT pyramid, SIDWT pyramid, morphological pyramid, proportion pyramid, differentiates pyramid, etc. All the above strategies, share one trademark: every technique is proficient for particular sorts.

MEDICAL IMAGE FUSION USING PCNN

The Medical imaging has become an integral part of modern medicine, its application throughout the clinical work, not only widely used in disease diagnosis but also plays an important role in surgery and radiation therapy program design, program implementation and efficacy assessments.

E.g. MR applicable to tumor tissue contour description, CT images can be accurately calculated tumor dose. PCNN is a naturally enlivened neural system in light of the work by Eckhorn in the execution of these calculations was finished by Johnson and his partners. They are neural models proposed by demonstrating a feline's visual cortex and created for elite biomimetic picture processing. In 1989, Eckhorn acquainted a neural model with copy the instrument of feline's visual cortex. The Eckhorn demonstrate gave a straightforward and powerful apparatus for concentrate less well-evolved creature's visual cortex and was soon perceived as having noteworthy application potential in picture processing [5]. It is characterized by the global coupling and pulse synchronization of neurons. These characteristics benefit image fusion which makes use of local image information.

PCNN has been successfully used in image fusion [3, 4]. In fact, humans are often sensitive to edges, directional features, etc. So, the pure use of single pixels is not enough. In 1994, Johnson adjusted the Eckhorn model to a picture handling calculation, calling this calculation a heartbeat coupled neural system. Over the previous decade, PCNNs have been utilized as a part of an assortment of picture handling applications, including picture division, highlight era, confronts extraction, movement identification, district developing, and commotion reduction. The essential property of the Eckhorn's connecting field show (LFM) is the coupling term. LFM is a tweak of the essential contribution by a one-sided balance, calculate driven by the connecting input. These drives an edge variable that rots from underlying high esteem. At the point when the edge dips below zero, it is reset to high esteem and the procedure begins once again. This is unique in relation to the standard coordinate and- fire neural model, which gathers the contribution until it passes a maximum point of confinement and viable "shorts out" to bring about the pulse. LFM utilizes this distinction to manage beat blasts, something the standard model does not do on a solitary neuronal level. It is profitable to see, nonetheless, that a point by point examination of the standard model must incorporate a shunting term, because of the coasting voltages level in the dendritic compartment(s), and thusly this causes an exquisite various tweak impact that empowers a genuine higher-arrange to organize (HON) [6].

Multidimensional heartbeat picture preparing of substance, structure information utilizing PCNN has been talked about by Kinser, et al. A PCNN is a two-dimensional neural system. Every neuron in the system compares to one pixel in info, picture, getting its relating pixel's shading data (e.g. Force) as an outer boost. Every neuron additionally associates with its neighboring neurons, accepting nearby jolts from them. The outside and neighborhood jolts are consolidated in an inside actuation framework, which aggregates the boosts until it surpasses a dynamic limit, bringing about a heartbeat yield. Through iterative calculation, PCNN neurons deliver a worldly arrangement of heartbeat yields. The fleeting arrangement of heartbeat yields contains data of information, pictures and can be utilized for different picture preparing applications, for example, picture division and highlight era. Contrasted and traditional picture preparing implies, PCNNs have a few huge benefits, including heartiness against commotion, freedom of geometric varieties in info designs, the capacity of crossing over minor force varieties in informative designs, and so forth. It has been demonstrated that PCNN is exceptionally reasonable for picture preparing, for example, picture division, picture upgrade, design acknowledgment, and so forth. Researchers have developed some image fusion algorithms based on PCNN however, all the image fusion methods using PCNN have one common trait: one PCNN cannot finish the whole process of image fusion.

Usually, a group of more than two PCNNs must be used to fuse multi-source images, making it inefficient and impractical, especially for a real-time system. Analysis of the PCNN exposes a defect preventing one PCNN from fusing multi-source images[13]. To make up for this defect, a new improved model, called m-PCNN, is proposed for the first time in this paper. Note: m indicates the number of external input channels. This model overcomes some limits of the original model in data fusion. A remarkable characteristic of m-

IIOAB JOURNAL
ISSN: 0976-3104

PCNN is that the number of external channels can easily be changed according to actual requirements, and is very useful when several images are fused at the same time. Therefore, m-PCNN successfully solves the problem of fusing multimodal images using only one PCNN. It has been proven here by experimental results that m-PCNN does well in the fusion of multimodal medical images. PCNN is a feedback network and each PCNN neuron consists of three parts: the receptive field, the modulation field, and the pulse generator.

In image processing, PCNN is a single layer pulse coupled neural cells with a two-dimensional connection. In the existing PCNN-based fusion algorithms, pixels in the spatial or MSD domain are input to PCNN, and there exists a one-to-one correspondence between the pixels and the neurons. Each neuron is connected with neighboring neurons in the linking range. The output of each neuron results in two states, namely firing and non-firing. Then, the sum of neuron firing times will generate a firing map whose size is equal to the images in spatial or MSD domain and the value of each pixel in firing map is equal to neuron firing times. We summarize these algorithms as. The value of pixels in the spatial or MSD domain is considered as the original image information in the existing algorithms. However, the pure use of pixels is not effective enough because humans are often sensitive to edges and directional features. We believe it will be more reasonable to employ features, rather than the value of pixels, to motivate PCNN. It is a self-organizing network that does not require training and the network was constructed by simulating the activities of the mammal's visual cortex neurons and the basic structure of the PCNN model. PCNN produces an output of binary pulse images when stimulated by images. Johnson and Padgett had enumerated the origin of PCNN, the basic model and the relation to biological models in their wide research on PCNN[7]. The number of neurons in the network is equal to the number of input images. Each pixel in the image is connected to a unique neuron and each neuron is connected with surrounding neurons through a radius of linking field.

The pulse coupled neural network has three compartments:

Receptive field

In this kind of attack, the attackers try to fetch the user's personal information without his permission. These are further of two types. The receptive field is the primary part to receive input signals from the neighboring neurons and from external sources and the field have two internal channels known as Feeding compartment F and linking compartment L. The linking inputs have faster characteristic response time constant when compared to feeding connections. The biased and multiplied linking inputs are multiplied with the feeding input to produce the total internal activity U which constitutes the Linking or Modulation part.

Modulation

In this kind of attack, the attackers try to fetch the user's personal information without his permission. These are further of two types. Receptive field is the primary part to receive input signals from the neighboring neurons and from external sources and the field have two internal channels known as Feeding compartment F and linking compartment L. Feeding input and Linking inputs communicates with the neighboring neurons through the synaptic weights M and W. Input stimulus is given only to the feeding compartment. The neuron is receiving the input stimulus S which is its corresponding pixel's color information along with the stimulus from neighbors in both the compartments.

Pulse generator

In this kind of attack, the attackers try to fetch the user's personal information without his permission. These are further of two types. The core reason why PCNN is used in image fusion lies in its global coupling and pulse synchronization of neurons. These biological characteristics make full use of the local information in images, but not single coefficient information in most popular MSD-based fusion algorithms. Although a regional firing characteristic of PCNN is investigated in multi-focus image fusion, we still use the firing times as a determination to select NSCT coefficients.

PCNN APPLICATIONS

- Image segmentation using PCNN Image segmentation is a technique that groups pixels into regions, and therefore defines object regions.
- PCNN dynamically evaluates the similarity between any two samples and this technique was utilized in segmenting fMRI images.
- Feature extraction- The intensities (color information), geometry structures (color distributions) of the images.
- Edge detection- Edge detection is based on hybrid harmony.
- Noise removal - Removal of extreme impulse noise from an image. It removes only the noisy pixels without disturbing the other pixels. Hence the image details and edge information could be preserved.

- Texture and fabric defects segmentation- The different gray intensity between the field of defects and the field with no defects was considered for PCNN neuron firing to implement segmentation.
- Surveillance techniques-The data processing with more number of synthetic aperture radar (SAR) images from spaceborne missions needs effective segmentation techniques.

SYNTHETIC APERTURE RADAR

Synthetic aperture radar (SAR (SAR) is broadly utilized as a part of the military observation, studying and mapping, direction, condition remote detecting and asset investigation. With the improvement of SAR, the exploration of programming target acknowledgment (ATR) has turned into an essential subject. SAR ATR is the way toward recognizing obscure focuses from its radar- resounded marks. On the off chance that the scale, position or the rise point of focus in the SAR picture transforms, it winds up hard to perceive the objective. The lucid idea of SAR symbolism prompts the presence of spot commotion in SAR picture and dot clamor causes a decrease of the picture quality and conceals the itemized structure of the picture. In this way, the SAR picture highlight extraction is one of the key difficulties of SAR programmed target acknowledgment. SAR ATR for ground-based targets is likewise critical in combat zone reconnaissance utilizing unmanned air vehicle-based frameworks. Among some effective SAR ATR procedures, guideline segment investigation (PCA) has been appearing to be a capable device for dimensionality lessening and highlight extraction [8].

A manufactured opening radar is an imaging radar mounted on a moving platform [17]. Electromagnetic waves are consecutively transmitted, and reflected echoes are gathered, digitized and put away from the radar receiving wire for later handling. As transmission and gathering happen at various times, they guide to various positions. The very much requested mix of the got signals constructs a virtual opening that is any longer than the physical receiving wire length. This is the reason it is named "engineered opening", giving it the property of being an imaging radar.[17] The range direction is parallel to flight track and perpendicular to azimuth direction, which is also known as along-track direction because it is in line with the position of the object within the antenna's field of view. The Moving and Stationary Target Acquisition and Recognition (MSTAR) is a joint Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL)effort to develop and evaluate advanced ATR system three BMP2 Armored Personnel Carriers (APC), a BTR70,2S1, BDRM2, D7, T62, ZIL131, ZSU23/4 [9].

CONCLUSION

The pulse coupled neural network is unique since each neuron of the PCNN represents an image pixel from the input image to be processed. The network is also powerful if the parameters are tuned properly. We have discussed the basic structure of the pulse coupled neural network and the consecutive changes made by the researchers. The survey proceeded towards the image processing techniques like image segmentation, image fusion, de-noising, feature extraction, pattern recognition, and miscellaneous applications. The network is compatible with all sorts of applications if modified. Based on the review done, the pulse coupled neural network is best suited for image processing applications. The flaws, information retrieval from the images, diagnosis of cracks or abnormalities from X-ray images, tumor detection from MR images, CT images, and removal of noise from images prove the importance of PCNN in the medical field. Generally, the widely used segmentation algorithms used for image processing has some common disadvantages like computational costs and more time consuming on due to certain features of images.

CONFLICT OF INTEREST

None.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the Accendere Knowledge Management Services.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Pacifici F, Frate FD. [2010] Automatic Change Detection in Very-High Resolution Images with Pulse-Coupled Neural Networks. IEEE Geoscience, and Remote Sensing Letters, 7:58– 62. doi: 10.1109/LGRS.2009.2021780

[2] Yue R, Huan YR. [2003] Pixel level fusion for multiple SAR images using PCA and wavelet transform. IEEE Transactions on Geoscience and Remote Sensing, 41(11):2540–2556

[3] Eckhorn R, Reitboeck HJ, Arndt M, Dicke P. [1990] Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex. Neural Computation. 2(3): 293 307.

[4] Xiao-Bo QU, Jing-Wen Y, Zi-Qian Z, Ben-Gang C. [2007] Multi-focus image fusion algorithm based on regional firing characteristic of pulse-coupled neural networks. In: Proceedings of International Conference on Bio-Inspired Computing. Theories and Applications. Zhengzhou, China: Publishing. House of Electronics Industry, 563-565. doi: 10.1109/BICTA.2007.4806419.

[5] Zhang Z, Blum RS. [1999] A categorization of multiscale decomposition based image fusion schemes with a performance study for a digital camera application. Proceedings of the IEEE, 87(8):1315-1326.

[6] Subashini M, Sahoo SK. [1999] Expert Systems with Applications, 41(20):3965–3974.

[7] Broussard RP, Rogers SK, Oxley ME, Tarr GL. Physio-logically motivated image fusion for object detection using a pulse coupled

neural network. IEEE Transactions on Neural Networks, 10(3): 554-563.

[8] Padgett J. [1999] IEEE Transactions on Neural Networks. for the shunting terms. C Lee Giles' old work from the late 80's on HONs), 10(3)480-498.

[9] Schumacher R, Rosenbach K. [2004] Atr of battlefield targets by sar classification results using the public star dataset compared with a dataset by qinetiq, uk, RTO SET Symposium on Target Identification and Recognition Using RF Systems. doi: 10.1109/INDICON.2014.7030432.

[10] Pacifici F, Frate FD, Solimini C, Emery WJ. [2007] An Innovative Neural-Net Method to Detect Temporal Changes in High-Resolution Optical Satellite Imagery. IEEE Transactions on Geoscience and Remote Sensing, 45:2940–2951. doi: 10.1109/TGRS.2007.902824.

[11] Bhatnagar G. [2010] Real time human visual system based framework for image fusion, in Proc. Member, IEEE, International Conference Signal and Image Processing, Trois-Rivieres, Quebec, Canada. 71–75. doi: 10.1007/978-3-642-13681-8_9.

[12] Jonathan Wu QM. [2012] Senior Member, IEEE, an image fusion framework based on the human visual system in the framelet domain. Int J Wavelets, Multires Inf Process, 10(1):21–30.

[13] Liu Z. [2010] Senior Member, IEEE, Real time human visual system based framework for image fusion, in Proc. International Conference Signal and Image Processing, Trois-Rivieres, Quebec, Canada, 75–78. doi: 10.1007/978-3-642-13681-8_9.

[14] Xiao-Bo QU. Master student in the Department of Communication Engineering at Xiamen University.

[15] Wald L. [1999] Some terms of reference in data fusion. IEEE Transactions on Geoscience and Remote Sensing, 37(3):1190–1193.

[16] Ramakrishnan S, Demarcus V, Le Ny J, Patwari N, Gussy J. Synthetic Aperture Radar Imaging Using Spectral Estimation Techniques. The University of Michigan.

[17] Jason M, Kinser A, Waldemark KB, Lindblad TB, Sven P. [2000] Jacobsson in Chemometrics and Intelligent Laboratory Systems. 51:115–124

[18] Mishra A, Sardar S. [2011] An improved algorithm for based imaging of breast tumors, in International Conference on Image Information Processing (ICIIP). doi: 10.1109/ICIIP.2011.6108901

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

NETWORK SECURITY USING LATTICE BASED CRYPTOGRAPHY

Sonal Singhla*, Shailja Yadav, Rohit Tanwar
Department of CSE, Manav Rachna Universitry,Faridabad, INDIA

ABSTRACT

In today's world, due to advancement of internet, our society becomes an information society and today era becomes an information era. Today's technology offered us variety of services with the spread of network, which comes with various threats to our confidential information, databases and website domains from cyber-attacks. So network security has become an issue which requires major attention. Lattice based cryptography has been a very active research area and still work is being done to design various cryptosystems that are efficient, easy to use and provide us with high degree of security This paper is an attempt to explain the network security and challenges and various techniques such as Lattice based cryptography through which network security can be enhanced.

KEY WORDS
Internet, Security, Cryptography, Encryption, Decryption, lattice.

Received: 27 Mar 2019
Accepted: 11 May 2019
Published: 10 June 2019

INTRODUCTION

Gestures and facial expressions can be used to communicate with the computers which require the computer system to understand and analyze the signals to perform a particular movement [1]. Recently the designing of special input devices proved to facilitate the interaction between humans and computers. Gesture recognition has been applied in a large range of application areas such as recognizing sign language, human computer interaction (HCI), robot control, smart surveillance, lie detection, visual environments manipulating, etc. Now a days different techniques and tools have been used for handling gesture recognition that vary between mathematical models like Hidden Markov Model (HMM) and Finite State Machine (FSM) to approaches based on software computing methods such as fuzzy clustering, Genetic Algorithms (GAs) and Artificial Neural Network (ANN). Since human hand is a complex articulated object which is controlled by 35 muscles and requires 27 degrees of freedom to be versatile in all the movements, it is a thrust area of research [2]. In today' digital field implementing gesture recognition system requires different type of devices such as cameras, instrumented gloves and colored markers.

Network is a group of two or more systems linked to each other to share resources such as files and information. A node is a point of connection within a network. It can be a computer or other devices such as scanner, printer, modem, etc. [3]. As the number of internet users is increasing rapidly, network security has become a challenge for network administrators to protect it from intruder's attack. These intruders attempt to obtain, alter or affect the original data at target machine. This results in undesired outputs and behavior. So the importance of network security becomes more significant.

Network security refers to the policies adopted for securing system resources from malicious programmers. All internet users, from an ordinary surfer to large enterprises require network security against the intruders [1]. Cryptography refers to the technique used to secure communication in the presence of third party using a public key i.e. hiding information by converting it into mystery code or cipher text. It is the branch of both mathematics and computer science. It includes encryption and decryption processes [4]. It is important to understand and learn the concept of network security for the following reasons. Following are the goals of network security [4].

- It maintains the privacy of an individual by denying the unauthorized access.
- It maintains the integrity i.e. assuring the consistency in data worldwide.
- It gives high availability.

CLASSIFICATION OF ATTACKS

Unethical practices in the field of IT have resulted into many problematic areas we need to tackle, such as different forms of attacks through which the intruder tries to affect the target machine.

We can classify the attacks as [1]-

Passive attacks

In this kind of attack, the attackers try to fetch the user's personal information without his permission. These are further of two types-

***Corresponding Author**
Email:
sonalsinghal1010@gmail.com

Traffic analysis

In this method, without the consent of the sender and the user, the third party (stalker) can view the hided information.

Release of message content

It is an easy approach to fetch the sender's and receiver's information.

Active attacks

It affects the system resources by altering the original data. It is classified as-

Modification of messages

Some messages are altered in order to produce an unauthorized effect.

Denial of services (DoS)

These services are a menace to the society as they require only some fundamental knowledge.

Distributed denial of services (DDoS)

Multiple affected systems are used to damage a single system.

Relay attack

In this kind of attack, intruder fetches the user's information such as credit card details via fake calls or doing some tricks.

Masquerade attack

The attacker gets the user's information such as login id and password by acting as some authorized party and then uses that information against them.

Insider attack

In such kind of attacks, an insider i.e. someone who belongs to authorized party invades the system resources in malicious ways and then damage the data on target machine.

Close-in attack

By applying some tricks, the hackers attack the target and fetch the information.

Phishing

The malicious hackers obtain the confidential information by fake calls or duplicate websites or mails pretending to be an authorized person.

Exploit attack

Intruders take the advantage of the system vulnerabilities.

Password attack

Intruders try to fetch user's password and get the confidential information.

Appearance based approaches

In this type of approach, visual appearance of input hand image is modelled using the feature extracted from stored image. Appearance based approaches are simpler and easier than 3D model based approaches due to the easier extraction of features in a 2D image. The common method used in this approach is to detect the skin colored regions in the image; however this method is affected by changing illumination conditions and other background objects with skin like color.

NETWORK SECURITY THREATS

These threats are increasing day-by-day rapidly and so the concern for this issue is increasing. Network security threats are classified into four categories as shown in [Fig. 1].

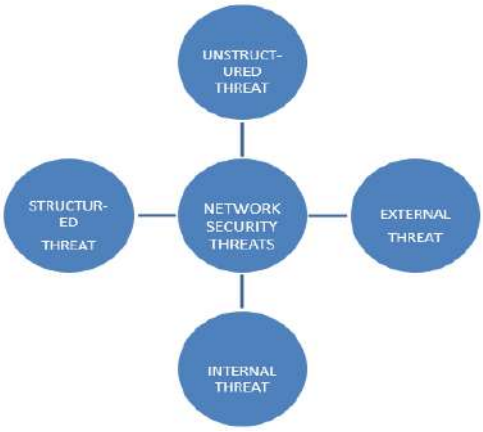

Fig. 1: Network Security Threats.

Unstructured threat

Unskilled individuals try to attack the target machine using some hacking tools in leisure time to do something challenging.

Structured threat

Malicious hackers who are technically skilled and are masters in this field generally hired by industries and organizations.

External threat

Threats from outsiders who do not belong to organization and attacks and steal the original data at target machine.

Internal threat

Threats from insiders who basically belong to the organization and still try to manipulate the original data by creating false streams resulting in false output.

TYPES OF NETWORK SECURITY

Network security threats can be classified in the following form as discussed in the [Fig. 2].

IMPLEMENTATION TOOLS

A lot of implementation, hardware and software tools have been used for recognizing gestures depending on the application fields they are used.

TEXANOMY

The taxonomy of the system can be analyzed using three factors: of the hand gesture application areas are mentioned below-

Encryption

It is a method of converting simple information (called plain text) into unintelligent (called cipher text).

Decryption

It is the reverse of the encryption which means cipher text into ordinary one.

Cryptosystem

It is a method of converting simple information (called plain text) into unintelligent (called cipher text). It is an organized list of attributes which contain all the finite possible planed and cypher texts. It also includes encryption and decryption algorithms [4].

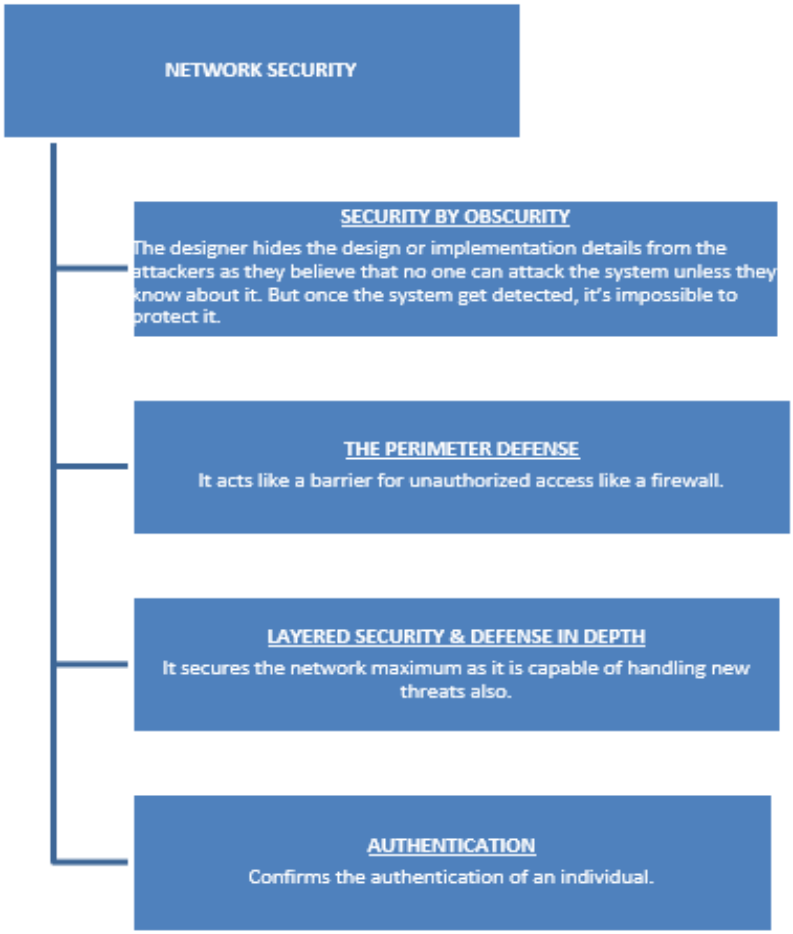

Fig. 2: Types of Network Security

SECURITY TECHNIQUES

The taxonomy of the system can be analyzed using three factors: of the hand gesture application areas are mentioned below- With the increase in internet users and technology, network security is more concerned. To enhance the network security, following techniques are used [3].

Hashing

A hash value is a number generated by a string of a text, also known as a message digest. The length of the hash value is generally less than the length of the input message. And it is practically impossible to create a duplicate text of the same hash value.

Symmetric key cryptography

It is also known as secret key cryptography. It uses the concept of a single key for both processes, i.e. encryption for plain text and decryption of unintelligible text.

Diffie-hellman key exchange (DH)

This technique was launched by Diffie and Martin Hellman in 1976. It allows the two parties to secretly create a secret key which can be used to communicate over insecure channel. Problem associated with it was that it was not practically possible.

Public key cryptography

It is a technique which includes two keys- public key and private key. The public key is known by both the parties while the private key is only recognized by the recipient. Public keys are used for the encryption of the information, and it can be decrypted only with the private key. But practically it was difficult to make the two keys to work simultaneously.

IIOAB JOURNAL
ISSN: 0976-3104

Elliptical curve cryptography

It uses smaller Abelian (finite) group i.e. subset of a lattice but its operations were costly.

Lattice based cryptography

It is widely used and details about this technique are discussed in brief.

LATTICE BASED CRYPTOGRAPHY

History of Lattice Based Cryptography [5-9]- In 1982, first time lattice was used in cryptanalysis. In 1995, Ajtai & Dwork described lattice based cryptography theoretically but they were unable to explain it practically. In 1996, Goldreich, Goldwasser & Halevi (GGH) introduced crypto schemes based on hard lattices, but it requires megabyte size public keys to secure the system. Later on smaller keys were used which helps in faster encryption. In 2009, encryption techniques based on lattices.

Lattice is a fundamental algebraic structure which consists of regular array of points in space. Lattice based cryptography is an efficient and easy implementation. It is the best way to secure quantum computers i.e.it provides security proofs based on worst conditions. When we talk about, lattice based cryptographic constructions, in terms of security; we divide it into two types.

- The first type focus on its practical aspects, but often lack security.
- The second type ensures a great proof of security, but only few are efficient and easy to use for practice.

A lattice is a discrete subgroup of R^n, or the set L (b_1, b_2... b_d) of all linear combinations $\sum x_i b_i$ where $x_i \in Z$, and the b_i's are linearly independent [5]. Now, if we talk about lattice crypto, it uses finite Abelian group. The Gaussian Heuristic theorem says that for full rank lattices L and uncountable sets C we have:

$$C \text{ and } (LC) = vol(C)/vol(L) \quad (1)$$

It basically measures the density of lattice. Hermite's Constant (1850): It is considered to be the worst case for short lattice vectors. From the last 20-25 years, lattices tend to be very trendy in case of complexity. It can be either classical or quantum.

Lattice based cryptography is highly in use in network security because of its primitives that are associated with problems such as CVP (closest vector problem) and SVP (shortest vector problem). In CVP, challenge is to find a closest point P. But it is difficult to find solution in higher dimension such as 500 [5]. In Lattice based cryptography, we come across various problems based on the presumed hardness. If we talk about SVP, there is no specifically an efficient algorithm to approximate SVP [6].

Let us suppose g=n^c in the worst case where, (n is the dimension of the lattice and c is the arbitrary constant independent of n), one can build knapsack-like cryptographic (one-way) functions that are almost certainly hard to break (when the key is chosen at random). Quantum computers can't even solve these problems [6].

Now, we will talk about its advantages like how it helps us other than lattice.

- More efficiency,
- It has security properties based on worst case assumptions e.g. quantum computers.
- Easy to design
- It has cheaper operations

CONCLUSION

As the internet is expanding with tremendous speed, the concern for security has been increased. It became the most attractive hunting ground for malicious hackers. These threats and outdated policies are one of the issues for major concern. And the hackers are finding more and more ways to penetrate the system resources and doing the malicious activities. So, in present scenario, there is a need for regular testing for network security [10]. Lattice based cryptography has been a very active research area and still work is being done to design various cryptosystems that are efficient, easy to use and provide us with high degree of security.

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
We would like to express our sincere gratitude to Accendere Knowledge Management Services., for providing us platform and opportunity to pursue the research.

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] Sumathi R, Sundarraja R. [2008] An Efficient Operator based Unicode cryptography Algorithm for Text, Audio and Video Files, 7(2): 1-14.
[2] Muley MN. [2015] Analysis for Exploring the Scope of Network Security Techniques in Different Era: A Study, IJACEN, 3(12): 33-36.
[3] Singh H. [2016] Network Security, a Challenge, IJARCCE, 5(3): 57-61.
[4] Kumar SN. [2014] Technique for security of multimedia using neutral network, Paper id- IJRETM-2014-02-05-020, IJRETM. 02(05):1-7.
[5] Peikert C. [2016] A Decade of Lattice Cryptography. Foundation and trends in Theoretical Computer Science, 10(4): 283-424.
[6] Micciancio D. [2003] Lattice Based Cryptography, Springer, 10(3): 147-191.
[7] Gama N, Nguyen PQ. [2008] Predicting lattice reduction. In Advances in Cryptology – Proc.Eurocrypt '08, Lecture Notes in Computer Science, Springer, 1-21.
[8] Silva R, Antonio C, de A, Campello JR, Dahab R. [2011] LWE-based identification schemes. In Information Theory Workshop (ITW). IEEE, 292–296. doi: 10.1109/ITW.2011.6089439
[9] Yao Y, et al. [2011] a Sub-0.5V Lattice-Based Public-Key Encryption Scheme for RFID Platforms in 130nm CMOS: 2-19. doi: 10.3233/978-1-60750-722-2-96.
[10] Lyubashevsky V. [2008] Lattice-Based Identification Schemes Secure Under Active Attacks. In Ronald Cramer, editor, Public Key Cryptography – PKC, number 4939 in Lecture Notes in Computer Science, Springer, 162–179. doi: 10.1007/978-3-540-78440-1_10.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

IMPLEMENTATION OF SCRUTINIZING SECURITY

Tridev Nath Tripathy*, Rajeev Singh

Department of CSE, Manav Rachna International Institute of Research & Studies, Faridabad, INDIA

ABSTRACT

Securing the cyber world has become the greatest challenge of all time. Cybercrimes are increasing uniformly all around the world. The undesirable acts are being done to gain the vital information of the well-known companies and individuals. My tool is a step towards the safer cyber world. It just scans the system and tell about the vulnerabilities of the system that can be used to exploit it and generating the alerts of the vulnerabilities of the system may be used to gain the access of the system.

KEY WORDS
Cyber Security, Armitage, Meta spolit Framework, VAPT, Information Gathering.

INTRODUCTION

At Security of the cyber world is required to enhance the privacy of an individual. On the daily basis, there are various companies that are affected by the cybercrimes. There are various steps taken by the organizations to protect their systems. The offenders gain the access to the system through two ways.

Hardware hacking

In the hardware hacking mostly there is an involvement of the pen drives or key loggers. These hardware's are enough to provide the important information of the system.

Software hacking

In the software hacking the felon enters into the system remotely by exploiting the services that are running on the victim's computer.

We have done the work to generate the report of vulnerabilities in the system by scanning and testing the exploits on it remotely. It is an automated tool for stepping forward in the direction of the world safer.

Received: 28 Mar 2019
Accepted: 21 May 2019
Published: 11 June 2019

LITERATURE REVIEW

After studying and going through various research papers we found that it is almost impossible to provide 100 percent security to the system. But we can defend our systems among the known attacks around the world, which makes very difficult for the attacker to enter into the system. The basic way to protect the systems are firewalls, then there are VAPT (Vulnerability Assessment and Penetration Testing) tests that are performed by the computer experts, but if the experts have to do the VAPT for a company then it requires a lot of experts if there is no automated system. All the experts have to do the VAPT of each and every computer one by one. So, by this project we try to solve this problem and automate the VAPT of the network.

ORGANIZATION OF THE PAPER

The paper is organized in the four divisions. In the first division we discuss about the previous related tools. In the second section we describe our project. In the third section, we discuss the difference. In the fourth section we give the conclusion and proposed work and in the fifth section we have given the references.

EXISTING SOFTWARE'S

There are various software's that are used for the VAPT of the systems for example:- Nmap, Metasploit Framework, Armitage.

***Corresponding Author**
Email:
tridev098@gmail.com

Nmap

It is used to scan the remote system, which helps us to gather information about the system like which ports are open and what service is running on them. It is also used to enumerate various types of services like smb service etc.

IIOAB JOURNAL
ISSN: 0976-3104

Metaslpoit framework

It is a framework used to exploit the various platforms. There are various exploits that are uploaded to its database and are used to gain excess of the system.

Armitage

It is a GUI form of the Nmap and Metasploit Framework.

PROJECT HORUS

One It is automated tool and can be used to do VAPT of the whole network. It has a four tier operation:

Pinging Network

Firstly, it scans the whole network and get the IP's that are alive and a list is formed [Fig-1].

Fig. 1: HORUS Welcome Message Network Security Threats

Scanning IP's

The second step involves the scanning and enumerating of the alive IP's also known as 'Information Gathering'.

Attack IP's

Based upon the information available it tries to exploit each and every combination that is possible.

Report

Based upon the information available it tries to exploit each and every combination that is possible. After attacking the IP's a report is generated which shows the vulnerabilities of the respective IP and if the IP exploitation is successful it generates high alert for that IP [Fig-1].

Fig. 2: Implementation of Scrutinizing Security

COMPARATIVE STUDY

After going through many research papers and tool reports, we concluded that there is a large difference between Horus and other tools. First and fore most difference is that there is no tool which is able to

perform the automated VAPT. Secondly, it is a able to provide better results in the network also. Thirdly, if we consider the scanning part Horus first version is weaker as compared to the other scanning tools.

CONCLUSIONS AND FUTURE WORK

In the first version of the tool we are able to perform simple scanning of networks and IP's and tries to exploit the systems by the familiar exploits. In the future we are trying to scan the IP's by various other methods and will add the feature of scanning the services that are running on the system.

CONFLICT OF INTEREST
None

ACKNOWLEDGEMENTS
We would like to express our sincere gratitude to Accendere Knowledge Management Services., for providing us a platform and opportunity to pursue the research.

FINANCIAL DISCLOSURE
None

REFERENCES

[1] www.wikipedia.org
[2] http://www.manchester.ac.uk/research/d.armitage/publications.
[3] https://nmap.org/bennieston-tutorial/.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

CLOUD COMPUTING DATA SECURITY USING ENCRYPTION ALGORITHMS

Kumar Rithvik[1], Simran Kaur[1], Shilpa Sejwal[1], Priti Narwal[1], Prateek Jain[2]*

[1]*Department of Computer Science & Engineering, Manav Rachna International Institute of Research & Studies, Faridabad, INDIA*

[2]*Accendere Knowledge Management Services Pvt. Ltd., INDIA*

ABSTRACT

The "cloud" is a set of different types of hardware and software that work collectively to deliver many aspects of computing to the end-user as an online service. Cloud Computing is the use of hardware and software to deliver a service over a network. With cloud computing, users can access files and use applications from any device that can access the Internet. There are a wide range of companies and industry verticals that use cloud computing such as Amazon and Google. While no storage solution is 100% safe, cloud storage providers can offer a safer and more accessible place for companies to store data than traditional computing methods. As Cloud computing has emerged as a new technology, it has also created new challenges such as data security, data ownership and trans-code data storage. In this paper we have discussed about cloud computing security issues, challenges that CSP (cloud service provider) face and study various security algorithms.

KEY WORDS
Algorithms: RSA, Diffie Hellman, DSA AES ,DES, Triple DES, Blowfish, Cloud Computing, Data Security

Received: 2 April 2019
Accepted: 26 May 2019
Published: 16 June 2019

***Corresponding Author**
Email:
Prateek.jain@accendere.co.in

INTRODUCTION

Cloud Computing is an important concept in computer development. It refers to the use of computing capacity, storage of computers and servers in the world over the Internet. It provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications [1]. Characteristics of cloud computing includes on-demand self-service, broad network access and resource pooling, rapid elasticity. Cloud systems automatically control and optimize resource use by a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts) [2]. Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service. Such cloud services are deployed in various ways which includes Software-as-a-Service (SaaS): which refers to software available on demand, it is based on multi-tenant architecture. Software like word processor, CRM (Customer Relation Management), etc. or application services like schedule, calendar, etc. are executed in the –cloudǁ using the interconnectivity of the internet to do manipulation on data. Custom services are combined with3rd party commercial services via Service oriented architecture to create new applications [3]. It is a software delivery for business applications like accounting, content delivery, Human resource management (HRM), Enterprise resource planning (ERP) etc on demand on pay-as-you go model. Another method to deploy cloud is Platform-as-a-Service (PaaS): This layer of cloud provides computing platform and solution stack as service [4]. Platform-as-a-Service provides the user with the freedom of application design, application development, testing, deployment and hosting as well as application services such as team collaboration, web service integration and database integration, security, scalability, storage, persistence, state management, application versioning, without thinking about the underlying hardware and software layers by providing facilities required for completion of project through web application and services[5].Cloud can also be deployed as Infrastructure-as-a-Service (IaaS): Infrastructure as a service delivers a platform virtualization environment as as service. Instead of purchasing servers, software, data centre space or network equipment, clients can buy these resources as outsourced service. In other words the client uses the third party infrastructure services to support its operations including hardware, storage, servers [6]. Private cloud is a new term that some vendors have recently used to describe offerings that emulate cloud computing on private networks. It is set up within an organization's internal enterprise datacenter [7]. Now the clouds can be of three types Private cloud, Public cloud and Hybrid cloud. In the private cloud, scalable resources and virtual applications provided by the cloud vendor are pooled together and available for cloud users to share and use. Only the organization and designated stakeholders may have access to operate on a specific Private cloud [8]. A public cloud is a model which allows users access to the services and infrastructure and are provided off-site over the Internet. It'stypically based on a pay-per-use model, similar to a prepaid electricity metering system which is flexible enough to cater for spikes in demand for cloud optimization [9]. Public clouds are managed by third parties or vendors over the Internet. Public clouds are less secure than the other cloud models because it places an additional burden of ensuring all applications and data accessed on the public cloud are not subjected to malicious attacks. However, security and governance issues must be well planned and ample security controls was put in place. A new concept combining resources from both internal and external providers will become the most popular choice for enterprises. A hybrid cloud is a combination of public and private cloud models that tries to address the limitations of each approach. In a hybrid cloud, part of the service infrastructure runs in private clouds while the remaining part runs in public clouds [10]. Hybrid clouds offer more flexibility than both public and private clouds. Specifically, they provide tighter control and security over application data compared to public clouds, while still facilitating on-demand service expansion and contraction. On the down side, designing a hybrid cloud requires carefully determining the best split between public and private cloud components [11].

IIOAB JOURNAL
ISSN: 0976-3104

Security in cloud computing involves concepts such as network security, equipment and control strategies deployed to protect data, applications and infrastructure associated with cloud computing. An important aspect of cloud is the notion of interconnection with various materials which makes it difficult and necessary securing these environments. Security issues in a cloud platform can lead to economic loss, also a bad reputation if the platform is oriented large public and are the cause behind the massive adoption of this new solution. The data stored in the cloud for customers represents vital information. This is why the use of such data by an unauthorized third party is unacceptable. There are two ways to attack data in Cloud [5]. One is outsider attack and the other is insider attack. The insider is an administrator who can have the possibility to hack the user's data. The insider attack is very difficult to be identified. So the users should be very careful while storing their data in cloud storage [9]. Hence, the need to think of methods that impede the use of data even though the data is accessed by the third party, he shouldn't get the actual data. So, all the data must be encrypted before it is transmitted to the cloud storage. Security allows the confidentiality, integrity, authenticity and availability of information [7]. The development of technologies and their standardization makes available a set of algorithms and protocols for responding to these issues [12].

The objective of this paper is to analyze data encryption algorithm for security purpose in cloud computing. The rest of the paper is organized in the following way. Section 2 deals with Symmetric Algorithms including DES, BLOWFISH, AES, Triple DES and Section 3 deals with Asymmetric Algorithms including RSA and Diffie-Hellman in section 3 of the paper.

RELATED WORK

A brief review of latest technology used for security in cloud computing has been presented. Thakur and Kumar provide a fair comparison between three most common symmetric key cryptography algorithms: DES, AES, and Blowfish. Since main concern here is the performance of algorithms under different settings, the presented comparison takes into consideration the behavior and the performance of the algorithm when different data loads are used. The comparison is made on the basis of these parameters: speed, block size, and key size [1].

Seth et al. performs comparative analysis of three algorithm; DES, AES an RSA considering certain parameters such as computation time, memory usages and output byte. A cryptographic tool is used for conducting experiments. Experiments results are given to analyses the effectiveness of each algorithm [14].

The comparative analysis of five algorithm; DES, 3DES, AES, UMARAM and UR5 Algorithm, considering certain parameters such as throughput, encryption time and power consumption has been performed by Pavithra, & Ramadevi [5]. A cryptographic tool is used for conducting experiments. The experimental results show the superiority of our UR5 encryption algorithm over other algorithms in terms of the power consumption, processing time, and throughput [4].

Shaina Arora et al. design an algorithm to merge both enhanced RSA algorithm and El-Gamal algorithm to provide user with a higher level of data security. The enhanced RSA algorithm enables faster encryption and decryption process and generating public and private key faster than the original RSA [8].

The implementation of Data Encryption Standard algorithm, which is one of the symmetric key cryptography algorithms. The m file DES. m is created and the two functions encrypt and decrypt () are called into this file. This m file DES. m gives the time required for encryption and decryption in seconds for the entered text [7].

The comparative study of various cryptographic algorithms like AES, DES, RSA, Blow Fish, Elliptic Curve, SHA and MD5 and give a proper direction to the users for use of proper algorithm for securing of data in cloud computing environment [12].MD5 algorithm takes least encryption time whereas, RSA takes largest encryption time.

The comparison between three symmetric key cryptographic techniques namely as DES, AES and Blowfish algorithms in terms of time and security by using image simulation. Based on the image files used and the experimental result it was concluded that Blowfish algorithm consumes least encryption time and DES consume maximum encryption time. We also observed that Decryption of Blowfish and AES algorithms is better than DES algorithm [9-11].

The conventional algorithms, based on their benefits and drawbacks has been discussed in next section. We additionally have in comparison the significance of each these cryptographic techniques. This paper also offers an appropriate future opportunity related to these cryptographic techniques.

Md. Alam Hossain et al. describes the basic characteristics (Key Length, Block size) of symmetric (AES, DES, 3DES, BLOWFISH, RC4), Asymmetric (RSA, DSA, Diffie-Hellman, El-Gamal, Paillier), Hashing (MD5, MD6, SHA, SHA256) algorithms. Also we implemented five well- known and widely used encrypt techniques like AES, DES, BLOWFISH, DES, RC4, RSA algorithms and compared their performance based on the analysis of their encryption and decryption time for different file sizes in the local system [7].

Kapoor et al. proposed a hybrid cryptographic technique for improving data security during network transmission is proposed and their implementation and results are reported. The proposed secure cryptographic technique promises to provide the highly secure cipher generation technique using the RSA, DES and SHA1 technique [11].

This paper discussed well-known cryptographic algorithms and also demonstrates the basic differences between the existing encryption techniques. Regardless of the mathematical theory behind an algorithm, the best algorithm are those that are well-known and well-documented because they are well- tested and well-studied.

EXISTING ALGORITHMS FOR SECURITY

To provide secure communication over the network, encryption algorithm plays a vital role. It is the fundamental tool for protecting the data. Encryption algorithm converts the data into scrambled form by using ―the keyll and only user have the key to decrypt the data. In Symmetric key encryption, only one key is used to encrypt and decrypt the data. Another technique is using asymmetric key encryption; two keys- private and public keys are used. Public key is used for encryption and private key is used for decryption.

Symmetric algorithms

This section deals with the symmetric algorithms. Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link [13].

Data encryption standard

The Data Encryption Standard (DES) is a block cipher. It was the first encryption standard to be recommended by NIST (National Institute of Standards and Technology). DES is the block cipher—an algorithm that takes a fixed-length string of plain text bits and transforms it through a series of complicated operations into another cipher text bit string of the same length. In the case of DES, the block size is 64 bits [14]. DES also uses a key to customize the transformation, so that decryption can supposedly only be performed by those who know the particular key used to encrypt. The key ostensibly consists of 64 bits; however, only 56 of these are actually used by the algorithm. Eight bits are used solely for checking parity, and are thereafter discarded. Hence the effective key length is 56 bits.

The advantages of DES lie on two facts:

- The use of 56-bit keys: 56-bit key is used in encryption, there are 256 possible keys. A brute force attack on such number of keys is impractical.
- The nature of algorithm: Cryptanalyst can perform cryptanalysis by exploiting the characteristic of
- DES algorithm but no one has succeeded in finding out the weakness.

Algorithm: functionDES_Encrypt (M, K) where M = (L, R) M IP(M)
For round 1 to 16 do Ki SK (K, round)
L Lxor F(R, Ki) swap(L, R) end swap(L, R) M IP-1 (M)
return M End

The weakness has been found in the design of the cipher:

- Two chosen input to an S-box can create the same output.
- The purpose of initial and final permutation is not clear.

Blowfish

This was developed in 1993 for the replacement of DES. It is one of the most common public algorithms provided by Bruce Schneier. As Blowfish is a symmetric algorithm, the same keys are used for decryption as well as encryption, the only difference is that the input to the encryption is plaintext, for decryption the input is cipher text. Blowfish is a variable length key, 64-bit block cipher. No attack is known to be successful against this because of the size of cipher text as it involves the whole plain text and convert the same text in cipher text. Blowfish encrypts 64-bit blocks with a variable length key of 128-448 bits [15-16]. According to Schneier, Blowfish was designed with the followings objectives in mind: fast- Blowfish encryption rate on 32-bit microprocessors is 26 clock cycles per byte, compact- Blowfish can execute in less than 5 kb memory, simple-Blowfish uses only primitive operation -s, such as addition, XOR and table look up, making its design and implementation simple, secure- Blowfish has a variable key length up to maximum of 448-bit long, making it both secure and flexible[17].

Blowfish suits applications where the key remains constant for a long time (e.g. Communications link encryption), but not where the key changes frequently (e.g. Packet Switching).

Algorithm:

Divide x into two 32-bit halves: xL, xR For i = 1to 16:
XL= XL XOR Pi
xR = F(XL) XOR xR
Swap XL and xR Next I
Swap XL and xR (Undo the last swap.) xR = xR XOR P17
xL = xL XOR P18
Recombine xL and xR

Blowfish is one of the fastest block ciphers in general use, except when changing keys. Each new key requires pre-processing equivalent to encrypting about 4 kilobytes of text, which is very slow compared to other block ciphers. This prevents its use in certain applications, but is not a problem in others, such as SplashID. In an application, it's actually a benefit especially the password-hashing method used in Open BSD uses an algorithm derived from Blowfish that makes use of the slow key schedule. Blowfish is not subject to any patents and is therefore freely available for anyone to use. This has contributed to its popularity in cryptographic software.

The disadvantages of Blowfish are it must get key to the person out of band specifically not through the unsecured transmission channel. Each pair of users' needs a unique, so as number of users increase, key management becomes complicated. For example, N(N-1)/2 keys required. Blowfish can't provide authentication and non-repudiation as two people have same key. It also has weakness in decryption process over other algorithms in terms of time consumption and serially in throughput.

Advanced encryption standard

AES (Advanced Encryption Standard), is the new encryption standard recommended by NIST to replace DES in 1997. Brute force attack is the only effective attack known against it, in which the attacker tries to test all the charactes combinations to unlock the encryption. Both AES and DES are block ciphers. It has variable key length of 128, 192, or 256 bits; default 256[18].

It encrypts data blocks of 128 bits in 10, 12 and 14 round depending on the key size. AES encryption is fast and flexible; it can be implemented on various platforms especially in small devices. Also, AES has been carefully tested for many security applications. This algorithm is fully open to public scrutiny and comment; this ensured a thorough, transparent analysis of the designs submitted. The criteria for choosing the AES algorithm includes security, cost, implementation [19].Some of the advantages of AES are:

- AES is more secure (it is less susceptible to cryptanalysis than 3DES).
- AES supports larger key sizes than 3DES's 112 or 168 bits.
- AES is faster in both hardware and software.
- AES's 128-bit block size makes it less open to attacks via the birthday problem than 3DES with its 64-bit block size.

Algorithm :

Cipher(byte() input, byte() output)
{ byte(4,4) State;
copy input() into State() AddRoundKey for (round = 1; round < Nr-1; ++round)
{SubBytesShiftRowsMixColumnsAddRoundKey }
SubBytesShiftRowsAddRoundKey copy State() to output()
}

Triple DES

In cryptography, Triple DES (3DES), is a symmetric-key block cipher, which applies the Data Encryption Standard (DES) cipher algorithm three times to each data block.

The original DES cipher's key size of 56 bits was generally sufficient when that algorithm was designed, but the availability of increasing computational power made brute-force attacks feasible. Triple DES provides a relatively simple method of by increasing the key size of DES to protect against such attacks, without the need to design a completely new block cipher algorithm.
Some of the advantages of 3DES are:

- AES in Galois/Counter Mode (GCM) is challenging to implement in software in a manner that is
- both performant and secure.
- 3DES is easy to implement (and accelerate) in both hardware and software.
- 3DES is ubiquitous: most systems, libraries, and protocols include support for it.

IIOAB JOURNAL
ISSN: 0976-3104

Algorithm:

A naive approach to increase strength of a block encryption algorithm with short key length (like DES) would be to use two keys (K1, K2) instead of one, and encrypt each block twice: EK2(EK1(plaintext)).
If the original key length is n bits, one would hope this scheme provides security equivalent to using key 2n bits long. Unfortunately, this approach is vulnerable to meet-in-the-middle attack: given a known plaintext pair (x, y), such that y = EK2(EK1(x)), one can recover the key pair (K1, K2) in ~2n steps, instead of ~22n steps one would expect from algorithm with 2n bits of key[20].

Therefore, Triple DES uses a "key bundle" that comprises three DES keys, K1, K2 and K3, each of 56 bits (excluding parity bits). The encryption algorithm is:

cipher text = EK3(DK2(EK1(plaintext)))

I.e., DES encrypt with K1, DES decrypt with K2, then DES encrypt with K3. Decryption is the reverse:

plaintext = DK1(EK2(DK3(cipher text)))

I.e., decrypt with K3, encrypt with K2, then decrypt with K1. Each triple encryption encrypts one block of 64 bits of data.

In each case the middle operation is the reverse of the first and last. This improves the strength of the algorithm when using keying option 2, and provides backward compatibility with DES with keying option.

Asymmetric algorithms

This section deals with the asymmetric algorithms. Asymmetric algorithms (public key algorithms) use different keys for encryption and decryption, and the decryption key cannot (practically) be derived from the encryption key. Asymmetric algorithms are important because they can be used for transmitting encryption keys or other data securely even when the parties have no opportunity to agree on a secret key in private.

RSA

The most common Public Key algorithm is RSA, named for its inventors Rivest, Shamir, and Adleman (RSA). RSA is basically an asymmetric encryption /decryption algorithm. It is asymmetric in the sense, that here public key distributed to all through which one can encrypt the message and private key which is used for decryption is kept secret and is not shared to everyone. How RSA is going to work in cloud environment is explained as: RSA algorithm is used to ensure the security of data in cloud computing. In RSA algorithm we have encrypted our data to provide security. The purpose of securing data is that only concerned and authorized users can access it[20-21]. After encryption data is stored in the cloud. So that when it is required then a request can be placed to cloud provider. Cloud provider authenticates the user and delivers the data to user. As RSA is a Block Cipher in which every message is mapped to an integer. In the proposed cloud environment, Public key is known to all, whereas Private Key known only to user who originally owns the data. Thus encryption is done by the cloud service provider and decryption is done by the cloud user or consumer. Once the data is encrypted with the Public key, it will be decrypted using the corresponding Private Key only.

Algorithm

Key Generation: KeyGen(p, q) Input: Two large primes – p, q Compute n = p .q
φ (n) = (p - 1)(q - 1)

Choose e such that gcd(e, φ (n)) = 1 Determine d such that e .d ≡ 1 mod φ (n) Key: public key = (e, n)
secret key= (d, n)

Encryption: c = me mod n where c is the cipher text and m is the plain text

There are advantages and disadvantages of RSA algorithm. The advantages include; RSA algorithm is safe and secure for its users through the use of complex mathematics. RSA algorithm is hard to crack since it involves factorization of prime numbers which are difficult to factorize. Moreover, RSA algorithm uses the public key to encrypt data and the key is known to everyone, therefore, it is easy to share the public key.

The disadvantages include; RSA algorithm can be very slow in cases where large data needs to be encrypted by the same computer. It requires a third party to verify the reliability of public keys. Data transferred through RSA algorithm could be compromised through middlemen who might temper with the public key system.

IIOAB JOURNAL
ISSN: 0976-3104

Diffie-hellman

The Diffie–Hellman key exchange scheme was first published by Whitfield Diffie and Martin Hellman in (1976) [22]. Diffie–Hellman key exchange is a specific method of exchanging cryptographic keys. This method allows two parties that have no prior knowledge of each other to jointly establish a shared secret key over aninsecure communications channel. This key can then be used to encrypt subsequent communications using a symmetric key cipher. The algorithm is itself limited to the exchange of keys. The Diffie–Hellman key exchange algorithm depends for its effectiveness on the difficulty of computing discrete logarithms [15-17]. Key exchange Algorithm Let us assume the A and B want to agree upon a key to be used for encryption / decrypting messages that would be exchanged between them.
The Diffie-Hellman key exchange algorithm works as follows:

Firstly, A and B agree on two large prime numbers n and g. These two integers need not be kept secret. A and B can use an insecure channel to agree on them.

A chooses another large random number x and calculates c such that c=g x mod n A sends the number c to B

B independently chooses another large random integer y and calculate d such that d=g y mod n B sends number d to A

A now compute the secrete key K1 as follows K1= d x mod n B now computes the secret key K2 as follows. K2=c y mod n

Some are advantages of Diffie-Hellman key exchange scheme are:

- The security factors with respect to the fact that solving the discrete logarithm is very challenging, and
- That the shared key (i.e. the secret) is never itself transmitted over the channel.

Some are advantages of Diffie-Hellman key exchange scheme are:

- The fact that there are expensive exponential operations involved, and the algorithm cannot be used to encrypt messages - it can be used for establishing a secret key only.
- There is also a lack of authentication.
- There is no identity of the parties involved in the exchange

Digital signature algorithm

The Digital Signature Algorithm (DSA) is a Federal Information Processing Standard for digital signatures. It was proposed by the National Institute of Standards and Technology (NIST) in August 1991 for use in their Digital Signature Standard (DSS) and adopted as FIPS 186 in 1993.[22-23] A digital signature algorithm (DSA) typically consists of three algorithms: A key generation algorithm that selects a private key uniformly at random from a set of possible private keys. The algorithm outputs the private key and a corresponding public key a signing algorithm that, given a massage and a private key, produces a signature. A signature verifying algorithm that, gives a massage, public key and a signature, either accept or reject the massages claim to authenticity [23].

Key generation

Select a prime q of 160 bits
Choose $0 \le t \le 8$, select $2511+64t < p < 2512+64t$ with q/p-1 Select g in Zp and a=g(p-1)/q mod p.$\alpha \neq 1$
Select $1 \le a \le q-1$, compute y= α b mod p. Public key (p,q, α ,y) private key a
Signing

Select a random integer k, $0 < k < q$.
Compute r= (α k mod p) mod q. Compute k-1 mod q.
compute s=k-1*(h(m)+ar) mod q. Signature=(r,s).

Verification :

Verify $0 < r < q$ and $0 < s < q$, if not, invalid. Compute w=s-1 mod q and h(m).
Compute u1=w*h(m) mod q. u2=r*w mod q. Compute v=(α u1 yu2 mod p)mod q.
Valid if v=r.

The following are the advantages of using digital signatures:

- Speed: Businesses no longer have to wait for paper documents to be sent by courier. Contracts are easily written, completed, and signed by all concerned parties in a little amount of time no matter how far the parties are geographically.

- Costs: Using postal or courier services for paper documents is much more expensive compared to using digital signatures on electronic documents.
- Security: The use of digital signatures and electronic documents reduces risks of documents being intercepted, read, destroyed, or altered while in transit.
- Authenticity: An electronic document signed with a digital signature can stand up in court just as well as any other signed paper document.

The following are the disadvantages of digital signatures:

- Expiry: Digital signatures, like all technological products, are highly dependent on the technology it is based on. In this era of fast technological advancements, many of these tech products have a short shelf life.
- Certificates: In order to effectively use digital signatures, both senders and recipients may have to buy digital certificates at a cost from trusted certification authorities.
- Software: To work with digital certificates, senders and recipients have to buy verification software at a cost.
- Law: In some states and countries, laws regarding cyber and technology-based issues are weak or even non-existent. Trading in such jurisdictions becomes very risky for those who use digitally signed electronic documents.

CONCLUSION

Cloud computing has many advantages over traditional storage system but there are still problems concerning security that need to be solved. Security is a major requirement in cloud computing while we talk about data storage If we can eliminate this security issue, the future is going to be on Cloud for large as well as small companies. In this paper, we have suggested some solutions that allow storage of data in an open cloud. Data security is ensured by our algorithms such as symmetric and asymmetric algorithms. The symmetric encryption technique and the asymmetric encryption technique are important in encryption of sensitive data. Encryption algorithms play an important role in data security on cloud and by comparison of different parameters used in algorithms, it has been found that AES algorithm uses least time to of execution. RSA consumes longest memory size and encryption time. Blowfish algorithm has least memory requirement. DES algorithm consumes least encryption time. Our future will be considering some problems related to existing security algorithms and implement a better version of DES, 3DES, AES, RSA, IDES, Blowfish.

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] Thakur J, Kumar N. [2011] DES, AES and Blowfish: Symmetric key cryptography algorithms simulation-based performance analysis. International journal of emerging technology and advanced engineering, 1(2):6-12.

[2] Harinath D, Murthy MR, Chitra B. [2015], Cryptographic methods & performance analysis of data encryption algorithms in network security. International Journal of Advanced Research in Computer Science and Software Engineering, 5(7):680-8.

[3] Manju RD. [2015] Sudesh Kumar. Analysis on Different Parameters of Encryption Algorithms for Information Security International Journal of Advanced Research in Computer Science and Software Engineering, 5(8):104-108.

[4] Adekanmbi OO, Omitola OO, Oyedare TR, Olatinwo SO. [2015] Performance Evaluation of Common Encryption Algorithms for Throughput and Energy Consumption of a Wireless System. Journal of advancement in engineering and technology, 3(1):1-8.

[5] Pavithra S, Ramadevi E. [2012] Study and performance analysis of cryptography algorithms. International Journal of Advanced Research in Computer Engineering & Technology, 1(5):82-86.

[6] Kashyap S, Madan N. [2015] A review on: Network security and cryptographic algorithm. International Journal of Advanced Research in Computer Science and Engineering, 5(4):1414-8.

[7] Hossain MA, Hossain MB, Uddin MS, Imtiaz SM. [2016] Performance Analysis of Different Cryptography Algorithms. International Journal of Advanced Research in Computer Science and Software Engineering, 6(3):40-52.

[8] Arora S. [2015] Enhancing Cryptographic Security using Novel Approach based on Enhanced-RSA and Elamal: Analysis and Comparison. International Journal of Computer Applications, 112(13):12-19.

[9] Kahate A. [2013] Cryptography and network security, Tata McGraw-Hill Education.

[10] Laser JS, Jain V. [2016] A Comparative Survey of various Cryptographic Techniques. International Research Journal of Engineering and Technology (IRJET), 3(03):11-17.

[11] Kapoor V, Yadav R. [2015] A Hybrid Cryptography Technique for Improving Network Security. International Journal of Computer Applications, ISSN: 0975–8887, 3(2):45-51.

[12] Wollinger T, Guajardo J, Paar C. [2003] Cryptography in embedded systems: An overview, proceedings of the Embedded World 2003 Exhibition and Conference. 735-744.

[13] Devi A, Sharma A, Rangra A. [2015] Performance Analysis of Symmetric Key Algorithms: Des, Aes and Blowfish for

Image Encryption and Decryption. International Journal of Engineering and Computer Science, 4(6):6-12.
[14] Seth SM, Mishra R. [2011] Comparative Analysis of Encryption Algorithms for Data Communication, 1(1):4-9.
[15] Dandalis A, Prasanna VK, Rolim JD. [2000] A comparative study of performance of AES final candidates using FPGAs, International workshop on cryptographic hardware and embedded systems Aug 17. Springer, Berlin, Heidelberg. 125-140.
[16] Spanos GA, Maples TB. [1995] Performance study of a selective encryption scheme for the security of networked, real-time video. In Computer Communications and Networks, 1995. Proceedings. Fourth International Conference on IEEE, 2-10.
[17] Elminaam DSA, Abdual-Kader HM, Hadhoud MM. [2010] Evaluating the performance of symmetric encryption algorithms. IJ Network Security, 10(3):216- 222.
[18] Mandal PC. [2012] Evaluation of performance of the Symmetric Key Algorithms: DES, 3DES, AES and Blowfish. Journal of Global Research in Computer Science, 3(8):67-70.
[19] Singh L, Bharti RK. [2013] Comparative performance analysis of cryptographic algorithms. International Journal of Advanced Research in Computer Science and Software Engineering, 3(11):43-52.
[20] Fujisaki E, Okamoto T. [1999] How to enhance the security of public-key encryption at minimum cost. In International Workshop on Public Key Cryptography. Springer Berlin Heidelberg, 53-68.
[21] Kaushik S, Singhal A. [2012] Network security using cryptographic techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 2(12):16-22.
[22] Kadam KG, Khairnar PV. [2015] Hybrid RSA-AES Encryption for Web Service. International Journal of Technical Research and Applications, 51-56.
[23] Nadeem A, Javed MY. [2005] A performance comparison of data encryption algorithms. In Information and communication technologies. ICICT. First international conference on IEEE, 84-89.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

GABVIE MODEL FOR SOFTWARE TESTING

Seema Sharma*, Shaveta Bhatia
Faculty of Computer Applications, MRIIRS, Faridabad, INDIA

ABSTRACT

Software testing is an immensely important topic, which helps us to create robust and effective software. Software testing means to compare the expected outcomes and real outcomes to ensure that the software product is free from error. The software testing is essential as the software bugs could be costly and risky. The testing by considering the code is referred to as the white box testing. There are many factors that should be consider before testing the software like statement coverage, decision coverage, branch coverage, condition coverage etc. This paper proposes a novel and effective model to carry out white box testing by finding the importance of a variable of the program code using Genetic Algorithms. The model is promising and would therefore help both the researchers and practitioners alike. The fitness function in Genetic Algorithm has been calculated mathematically which outperforms the earlier traditional methods. It includes the information necessary for ascertaining the importance of a variable.

KEY WORDS
Software testing, Genetic Algorithm, Fitness Function, White box testing.

INTRODUCTION

Software testing has been conducted with an intention to find bugs and verify whether software is fit for use or not. It provides the information of quality of software to the stakeholders. Software testing is conducted for the analysis to check.

1. Whether it meets all requirements to design and development of the product
2. All types of inputs must be responded or not
3. All functions perform well or not

Software testing is very complicated task. It accounts for the major portion of the project time [1]. It can be classified as White box and Black box testing [3]. White box testing explores the code while the black box testing only sees the input and output [4]. The grey box testing lies in between them. Though many techniques are available for the efficient and effective testing of a program, as per the literature review, they do not consider the occupancy of a variable in a program [2]. This research has taken the above limitation by adding the parameters which are "variables" used in code for appropriate testing taken as the initial population in GA.

In this paper; a novel model is proposed to accomplish white box testing. The work uses the frequency of a variable used in the program code by calculating its dynamic priority based on its occurrence [2]. Furthermore, the fitness of a variable is calculated mathematically based on the rate of recurrence. Since the search space can become very large, therefore a heuristic search technique is needed to access the importance and hence to select a variable. A heuristic technique is fast calculation method for getting the optimal solution. Genetic Algorithm is a heuristic search process which is based on the survival of the fittest [19-22]. The explanation of the algorithm has been included in the paper.

Received: 26 Sept 2019
Accepted: 15 Oct 2019
Published: 20 Oct 2019

The work is organized as follows. Section 2 of this paper presents a brief literature review. Section 3 presents a discussion on Genetic Algorithms. Section 4 presents the proposed work and algorithm and the last section concludes.

LITERATURE REVIEW

There has been immense work has been done on software testing by various researchers. But still to develop the better model for testing the software. The proposed work deals with developing the model used to test the program code which is discussed in the next section. Some of the related works done by the prominent researchers and the issues in this field are discussed below. A new strategy has been proposed on the basis of existing work.

The combined features of Cuckoo search and genetic algorithm has been proposed by the authors Khan et al. in 2018 [14] for optimization of test cases. A cuckoo search algorithm first ameliorates arbitrarily generated test cases and after that, genetic algorithm is used to create new test cases. There is a need of more improvement in their algorithm in order to get efficient and optimized result. Also the cuckoo search algorithm is applied only for the amelioration of test cases and genetic algorithm is applied for the new generation of test cases, which increase the complexity of the algorithm. Bashir and Nadeem in 2017[10] proposed a technique for generating the test case for mutation testing. Since mutation testing is time consuming and expensive task therefore the author has taken genetic algorithm approach for reducing the cost. A new fitness function was introduced; eMujava and the results were compared with standard fitness functions. The authors have taken four parameters for testing fitness function for state based and object oriented program, two ways cross over and adaptable mutation. However, other versions of cross

***Corresponding Author**
Email:
seema.fca@mriu.edu.in
Tel.: +91-9873553830

over and mutation have not been taken.

A new approach have been implemented for the automatic test case generation in order to improve the efficiency of software by Khan and Amjad (2015) [13]. In this paper the author suggested that random method is inadequate for the selection process of test data. For this purpose the author used mutation analysis and genetic algorithm approach to produce test data for program code. The inter procedural control flow graph (CFG) is used for the program. This weight of CFG is distributed among all edges according to the pareto principle. However the algorithm introduced is suitable for unit testing but if we want to test the overall code then mutation testing is not effective. The Markov Chain approach with genetic algorithm has been applied by the authors Boopathi et al. 2014 [11] for testing the software. In his approach, the code is firstly converted into control flow graph and then optimized to implement DD-graph. The fitness function is based on coverage of path in DD-graph. The weights of edges are assigned with the help of Markov transition probability matrix and the fitness function of genetic algorithm is calculated as the sum of all the weights of edges in DD graph. The operators of genetic algorithm are applied for generating the test cases to cover maximum path. Only one point crossover and uniform mutation genetic operator is used. The results are not compared with simple genetic algorithm or genetic algorithm with varying population or with random techniques.

Saini and Tyagi (2014) [5] have used two different optimization techniques, Genetic Algorithm (GA) and Clonal Selection Algorithm (CSA)for test case generation. This paper presents to initiate the test data by using these techniques to test the basis path testing. Korel's Distance Function (Korel 1990) is used for accomplishment of fitness function. Since basis path testing means every statement and every branch should be test. But fitness function used examined only branch predicates. Various experiments have been evaluated in Matlab like square root, quadratic equation, trigonometry function, Linear equation etc and compare the results with the random-testing technique. Also using these techniques we cannot conclude, which algorithm is best GA or CSA. Minj and Belchanden (2013) [3] had presented the technique to produce test cases through UML State diagram, instead of control flow graph, which is totally based on path oriented approach. UML State chart modeling is used to show the control and flow of data. It comprises of five following parameters.

- Initial state
- Final state
- No. of intermediate States
- Transition function
- Process function

Nirpal and Kale (2011[4]) presents use of genetic algorithms in software testing for automatically creation of test cases. In this method, the target path is selected and then sequences of operators of genetic algorithm are executed in order to achieve the good test cases. They have used triangle classification program for experiment and compare it with Young Chen method method [12]. But if the code consists of many functions, the probability of testing the function using above approach may not work. Fitness function is evaluated by calculating distance covered between the executed path and the target path. Doungsa et al. [23] used the genetic algorithm for generating the test cases in path testing from UML state machine. The test data covered the maximum transition of the possible path. The number of test data generated was increased in number as one test data test only one path, so for maximum coverage the author had increased the number of test cases. However the result is suitable for the simple program without loop.

Our work proposed considers the variable found in the code as a parameter as well as all of the below parameters and promises to give much better performance for testing the program code

Table 1: Parameters used in Software Testing (Source-Self)

Type	Author Name	Cross Over rate	Mutation Rate	Selection	Population	Iterations	Encoding
Path testing	Rijwan Khan, Mohd Amjad and Akhlesh Kumar Srivastava 2018	0.8	0.2	Random	20	10	Binary
Path testing	Bashir, M.B. and Nadeem, A., 2017	0.5	[1-0]	Random	50	10	Real

Mutation testing	Rijwan Khan, Mohd. Amjad	Condition: Mutation score > 0.5	Condition: Mutation score > 0.5	Tournament	_	20	Real
Data Flow	M. Boopathi1, R. Sujatha, C. Senthil Kumar, S. Narasimman	0.6-0.8	0.0-0.2	Random	5	5	Real
Path testing	Poonam Saini , Sanjay Tyagi	0.8	0.15	Roulette wheel	20	100	Binary
Path testing	Jasmine Minj Lekhraj Belchanden:2013	0.8	0.2	Roulette whell method	4	10	Binary
Path testing	Premal B. Nirpal and K. V. Kale(2011)	0.9	0.01	Random	1-1000	1000	Binary
Parallel Path testing	Doungsa, Dahal, Hossain & Suwannasart	0.5	0.05	Random	10*10	100	Binary

MATERIALS AND METHODS

Genetic Algorithm (GA) is a heuristic search process based on the survival of the fittest [1]. The algorithm was invented by Holland in 1960's [2]. The algorithm is very useful for finding the solutions of optimization problems. The search proceeds in the following way. The process starts with an initial population. The population is composed of chromosomes and each chromosome has cells. The cell may depict the inclusion or exclusion of a particular feature, in case of a Binary population. The number of chromosomes is initially, low. The process builds up a buffer of chromosome; each generation is more fit, on an average as compared to the previous generation [16]. This is achieved by the crossover and mutation operator. The crossover operator takes two chromosomes and produces a new chromosome. The operator can use one point mechanism or two point crossover or multipoint cross over, from amongst the many available options. In the one point cross over, one crossover point is selected, binary string from beginning of chromosome to the crossover point is copied from one parent, the rest is copied from the second parent. In the case of two point cross over two points are selected and new chromosome(s) is produced. The mutation operator changes an existing chromosome and may end up producing a very good chromosome. As per the literature review the mutation operator inculcates diversity. Flip bit is one of the types of mutation. The above two operators are based on the selection of chromosomes. There are many types of selections like Rowlett wheel selection or tournament selection, to name a few. The Rowlett wheel selection selects a chromosome based on the commutative fitness. [17]

A chromosome is assigned a fitness based on the given problem. The fitness should contain the crux of the problem and can use functions like $f(x) = 1/(1 + e^{\lambda x})$, where λ can be found by empirical analysis. Note that the value of this function is between 0 and 1 and hence this function assigns a value which can be perceived as the probability of the chromosome being fit [18]. The termination of GA can be done either when the number of generations exceeds the set limit or there is practically no change in the chromosomes. The steps of Genetic Algorithm are depicted in [Fig. 1].

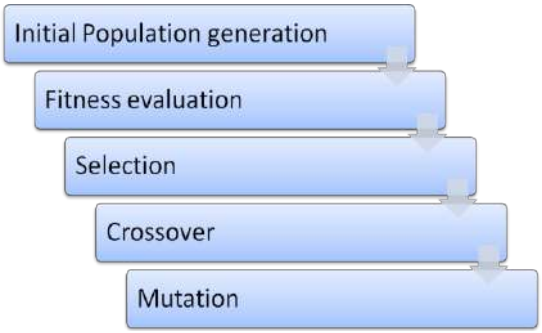

Fig. 1: Steps of each iteration of GA

The pivot concept behind the Genetic Algorithm based variable Importance evaluation (GABVIE) model is taking "variables" used in the program code as a constraint used to generate the test cases to test the path testing of a program code. The proposed algorithm assigns fitness to a variable depending upon its occurrence in the code and the frequency. The frequency of a variable is counted as follows. The occurrence of a variable in the 'if' block, assigns 1 to the frequency and that inside a 'for' loop (or 'while' for that matter) assigns n to it. The weight is determined by the frequency of the selected variables, multiplied by a constant λ, determined empirically. The above procedure would result in the selection of

the most important variables.

$$\text{Fittness} = 1/(1 + e^{\Sigma v_i w_i})$$

From amongst the given fitness function $1/(1 + e^{f})$ is chosen as it results in a value between 0 and 1, which can be perceived as the probability of the selection of variable. Here, v_i is 0 or 1, depending upon if the ith variable is selected or not and w_i is the weight. This is followed by the application of the GA process. The cross over and mutation applied till the population reaches its termination condition. The following algorithm and [Fig. 2] depicts the process.

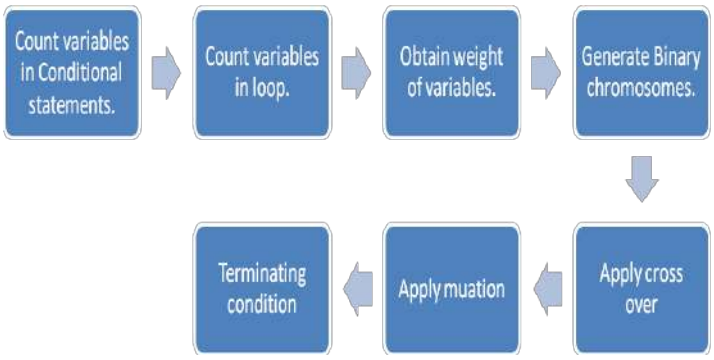

Fig. 2: Working model of GABVIE

Initially two arrays var and count are initialized to NULL. This is followed by populating the var[] array to variables in the program after scanning the program (Step 3). The program is again scanned for any occurrence of 'if' after which the count of the variable found is incremented by 1 and for 'loop' the count is incremented by 'm' the number of times, a loop runs. This is followed by the calculation of fitness. A binary population is then created. Row wheel selection is then applied to find the points on which one point cross-over is to be applied. Random mutation is then applied. The process is repeated till the stated numbers of generations are reached.

ALGORITHM:

Our key idea is to take a program code as a input and all the variables used in the program code are passed as parameters in the above model and cast this software testing problem as one of finding appropriate fitness of the functions in the program.

```
Algorithm 1: GABVIE v, t)
Input: Program statements with variables as parameters
Output: Fitness of the functions defined in program code
For i – 1 to n do
Scan
if equals (t, v[])
        then token <- var[ i++]
end if

Scan
If equals (token, block(if))
        Scan block(if)
        For i – 1 to n
                if equals (token, var [])
                count [] <- count[] +1
                end if
        end for
end if

if equals ( token, block(loop))
        Scan block(loop)
        For i – 1 to m
                if equals (token, var[])
                count <- count +m
                end if
        end for
end if

cell[]  <- var[]
```

```
create PopulationChromosome ( count, cell[])
for each cell
        if equals (cell [] , 1)
                calculate fitnessfunction
                fitness <-fit (count [])
                return fitness
        end if
end for

Select MostfittedChromosome
Point <-RowlettWheel (fitness, PopulationChromosome)
Calculate OnepointCrossover (point)
Calculate Mutation (random (PopulationChromosome))
Repeat until gen <=maxgen
```

The model has been implemented in Python and checked for few programs. It was observed that some conditions will be imposed on the input programs. In order to obtain the requisite fitness, the program has been divided in to two parts. The first part counts the occurrences of variables and decides the input to the fitness function and second part gives this input to the module implementing genetic algorithms.

[Table 2] shows the function used in the first in the first model and [Fig. 3] shows flow of the program.

Table 2: Functions used in model for Testing (Source-Self)

Function Name	Arguments	Description
_file	File Name	Opens code file and breaks it into a stream of tokens, returns the stream of tokens in a list named 'str'.
_var	str	Scans the input stream of code and selects the variables of data type int, char and float except for 'i' and 'j' which are reserved for loop iterations. Returns a list named 'var' containing the names of variables used.
_scan	str	Scans the code and detects the variables that are scanned by the user and takes the input by user. Returns a list 'scan_var' with the names of scanned variable and list 'value' with the respective value of those scanned variables.
_cou	temp, scan_var, value	'temp' contains the for loop condition; This function calculates the number of times a loop iterates. Returns the count of iterations evaluated by the expression in 'temp' as variable 'n'.
_count	str, var, scan_var, value,	Scans the code and calculates the number times a variable is used in the program. Returns a list 'count' containing the count of variables in 'var'

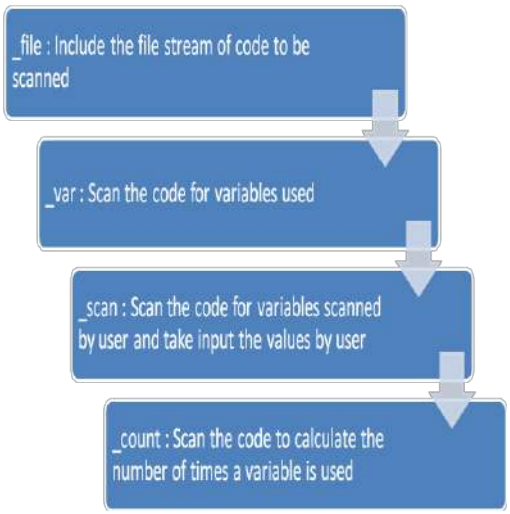

Fig. 3: Flow of Program

RESULTS

The model was tested the 10 C programs and it was found that the variables which were used more got the higher fitness value. It was also observed that the existing models cited above did not give requisite importance to the variables frequently used and hence was taking in the appropriate treatment of such variables leading to the test cases that may not contain such variables. In our case such variables were

being always a part of list being generated for creating test cases.

CONCLUSION

The work presented a novel model for selecting variables for testing. The model uses Genetic Algorithm. The fitness of the variables is determined using the frequency of occurrence. This is followed by the application of standard GA. The model has been carried out some results have been stated. The results are encouraging. Currently the model is being tested on a larger set of programs. The calculation of constants stated in the work would be done by extensive empirical analysis. The model has been proposed and the implementation being carried out. This model would pave way for the applicability of GA in testing.

ABOUT AUTHORS

Seema Sharma is a Research Scholar and Shaveta Bhatia is an Associate Professor, MRIIRS, Faculty of Computer Applications, Faridabad, India

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Alzabidi M, Kumar A, Shaligram AD. [2009] Automatic Software structural testing by using Evolutionary Algorithms for test data generations. International Journal of Computer Science and Network Security, 9(4):390-395.

[2] Girgis MR. [2005] Automatic Test Data Generation for Data Flow Testing Using a Genetic Algorithm J UCS. 11(6): 898-915.

[3] Minj J, Belchanden L. [2013] Path Oriented Test Case Generation for UML State Diagram using Genetic Algorithm. International Journal of Computer Applications, 82(7).

[4] Nirpal PB, Kale KV. [2011] Comparison of software test data for automatic path coverage using genetic algorithm development, 4(5).

[5] Saini P, Tyagi S. [2012] Test data generation for basis path testing using genetic algorithm and clonal selection algorithm, Int J Sci Res, 3(6):2319-7064.

[6] Sharma A, Rishon P, Aggarwal A. [2016] Software testing using genetic algorithms Int J Comput Sci Eng Surv (IJCSES), 7(2):21-33.

[7] Srivastava PR, Kim TH. [2009] Application of genetic algorithm in software testing. International Journal of software Engineering and its Applications, 3(4):87-96.

[8] Goldberg DE. [1999] Genetic and evolutionary algorithms come of age. Communications of the ACM, 37(3):113-120.

[9] Chauhan N. [2010] Software Testing: Principles and Practices. Oxford university press.

[10] Bashir MB, Nadeem A. [2017] Improved genetic algorithm to reduce mutation testing cost, IEEE Access, 5:3657-3674.

[11] Boopathi M, Sujatha R, Kumar CS, Narasimman S. [2014] The mathematics of software testing using genetic algorithm. In Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization IEEE, 1-6.

[12] Chen Y, Zhong Y, Shi T, Liu J. [2009]. August Comparison of two fitness functions for GA-based path-oriented test data generation. In 2009 Fifth International Conference on Natural Computation IEEE, 4:177-181.

[13] Khan R, Amjad M. [2015], December, Automatic test case generation for unit software testing using genetic algorithm and mutation analysis. In 2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON), IEEE, 1-5.

[14] Khan R, Amjad M, Srivastava AK. [2018] Optimization of Automatic Test Case Generation with Cuckoo Search and Genetic Algorithm Approaches. In Advances in Computer and Computational Sciences, Springer, Singapore, 413-423.

[15] Last M, Eyal S, Kandel A. [2005] Effective black-box testing with genetic algorithms. In Haifa Verification Conference Springer, Berlin, Heidelberg, 134-148.

[16] Mishra DB, Mishra R, Acharya AA, Das KN. [2019] Test Data Generation for Mutation Testing Using Genetic Algorithm. In Soft Computing for Problem Solving, Springer, Singapore, 857-867.

[17] De ABT, Martins E, De SFL. [2007] Generalized extremal optimization: an attractive alternative for test data generation. In Proceedings of the 9th annual conference on Genetic and evolutionary computation, ACM, 1138-1138.

[18] Knowles J, Corne, D. [1999] The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation. In Congress on Evolutionary Computation (CEC99), 1:98-105.

[19] Bhasin H, Singla, N. [2012] Harnessing cellular automata and genetic algorithms to solve travelling salesman problem. In International Conference on Information, Computing and Telecommunications, (ICICT-2012), 72-77.

[20] Blanco R, Tuya J, Adenso-Díaz B. [2009] Automated test data generation using a scatter search approach. Information and Software Technology, 51(4):708-720.

[21] Deb K, Pratap A, Agarwal S, Meyarivan TAMT. [2002] A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2):182-197.

[22] Engström E, Runeson P, Skoglund M. [2010] A systematic review on regression test selection techniques. Information and Software Technology, 52(1):14-30.

[23] Doungsa-ard C, Dahal KP, Hossain MA, Suwannasart T. [2007] An automatic test data generation from UML state diagram using genetic algorithm. In: Proceedings of the Second International Conference on Software Engineering Advances (ICSEA 2007). 25-31 Aug. 2007 Cap Esterel, France. IEEE Computer Society Press. 47-52.

ISSN: 0976-3104
ISSUE: COMPUTER SCIENCE

ARTICLE

ALGORITHMS TO ENHANCE THE RELIABILITY OF VIRTUAL NODES USING ADAPTIVE FAULT TOLERANCE TECHNIQUES

Mridula Dhingra*, Neha Gupta

Faculty of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad, INDIA

ABSTRACT

Cloud computing has various benefits in terms of cost saving, flexibility, efficiency etc. but it is not much efficient to perform real time computing in cloud infrastructure. The real time system acquires benefits of scalable virtualized environment and intensive cloud computing capabilities for executing real time tasks. Most of the real time applications undergo its processing on remote cloud computing nodes and chances of error occurrence is also high due to undecided latency. And it is essential that the safety critical applications should have higher reliability. To determine whether the fault tolerance mechanism is having the higher reliability and availability or not, an efficient adaptive fault tolerance mechanism has been proposed through reliability assessment architecture to enhance the reliability of the system. Also fine grained check point algorithm is used to minimize the latency.

KEY WORDS
Cloud Computing, Virtualization, Fault Tolerance, Reliability Assessment, Fine grained checkpointing.

INTRODUCTION

Now a days Cloud technologies have become a vital development in the era of information technology. Like every technology, cloud computing undergoes from some severe issues like fault tolerance of real time systems [1]. Virtualization and Internet-based Cloud computing leads to different kinds of failures to arise and hence, leads to the requirement for reliability and availability has turn out to be an essential issue. To ensure reliability and availability of cloud technologies, a scheme for fault tolerance need to be developed and applied. Majority of the early schemes for fault tolerance were paying attention on only one way to tolerate faults. Now cloud computing has changed the trend that how computing will perform, But the users are still adamant to work for real time applications. But many researchers are now working on real time applications. Real time systems are scrutinized on the basis of timeliness and fault tolerance factors[2]. Timeliness is the capability of a real time system to complete the proposed assignment within a specific time and fault tolerance is the ability of the system to work elegantly in the occurrence of fault. Now a days as the requirement of cloud computing is growing day by day, real time systems can be carried out on cloud environment. In most of the real time cloud applications, execution is done on remote cloud computing nodes. So the probability of errors become high because of undecided latency of computing node but on the contrary most of real time systems are safe and extremely reliable and to conquer the reliability of real time systems in cloud computing, there is huge demand of fault tolerance amongst users. [3] Cloud infrastructure has established many issues related to real time computing. In cloud computing, the latency of virtual nodes is not known. Even if one resolves the latency, but it amends after some period of time. In this research paper, this constraint has been overcome by proposing a framework based on adaptive reliability assessment and fault tolerance of VMs in real time applications.

Received: 3 Aug 2019
Accepted: 15 Oct 2019
Published: 20 Oct 2019

To attain reliability the cloud computing, vendors implements a variety of fault tolerance methods to run the system in case of faulty conditions [2].

Need of Fault Tolerance Methodology

- Fault tolerance provides the method or module to do the task effortlessly even though in the incidence of malfunction.
- Fault tolerance assures quality of service for users by definite recital.
- User can rearrange and append software instructions in accordance with their needs.
- Fault tolerance is a technique or concept to implement a system to perform well in unexpected situations, e.g. when a part of cloud is not working up to the level of satisfaction the whole server of the cloud will not collapse.

RELATED WORK

Limam and Belalem (2014) [4] proposed an adaptive checkpoint method in which the methodology to eliminate preventable checkpoints and adding of additional checkpoints was suggested. This methodology works on the current status of various cloud components. This method either enhances or reduces the intervals between checkpoints.

Cao et al (2015) [5] have presented a scheme to support energy jobs and jobs with priorities. This method is also based on checkpointing.

***Corresponding Author**
Email:
mridula.fca@mriu.edu.in
Tel.: +91-9871462056

Bobroff et al (2007) [6] have discussed the methods to forecast the workload of VMs and based on this workload classification of VMs is done. Authors have proposed a VM management algorithm which reduces the total power consumption on run time without compromising the SLA violation rates.

Verrna et al (2008) [7] have proposed a model based on VNs. Their methodology works on the assumption that no multiple copies of VM exist. The proposed method is based on a power-aware VM placement controller to control the heterogeneous servers and VMs.

Proposed work

In [Fig. 1], the proposed reliability assessment architecture is discussed. The proposed architecture will consist of N virtual nodes. Each node will take the input data from the input buffer and then the input will be given to the each virtual machine. Every time the input will be taken by the node, it will perform the operation and provide the result. The proposed reliability framework will primarily divided into three modules, these are:

- Acceptance Computation
- Reliability Assessment
- Fine-grained checkpointing
- Decision Mechanism
- Replication

Decision Mechanism and Replication will be discussed in next upcoming research papers.

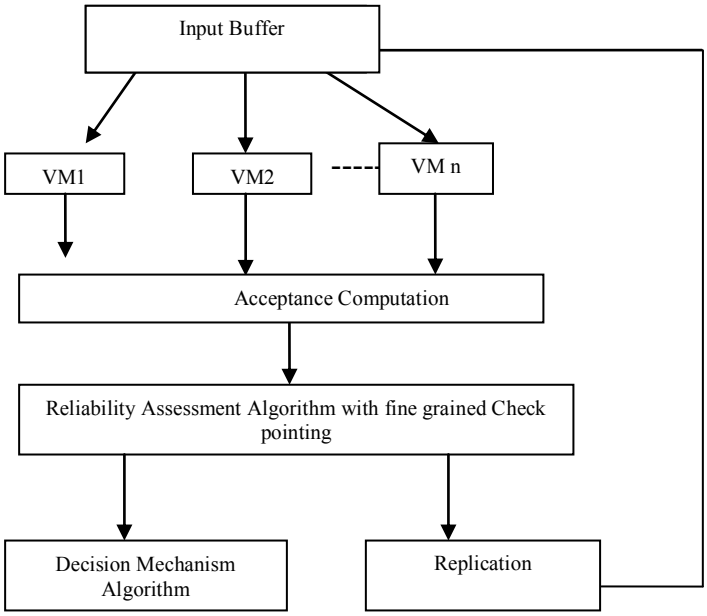

Fig. 1: Reliability assessment architecture.

Acceptance computation (AC): The proposed scheme will utilize n Virtual Machines wherein each machine acquires its input from input buffer. All virtual machines concurrently acquires input from it. Every single machine takes the input, executes the application and produces result. Then the results are directed towards the Acceptance Computation module. If the acquired results are correct, then its computational time is compared with threshold time. If the computational time exceeds threshold time, the node is considered as a failure node and its reliability is not tested and backward recovery will be performed through fine-grained checkpointing, otherwise the results will passed to reliability assessment module to measure the reliability of virtual machines.

Reliability assessment (RA): The reliability is a continuous measure and changes its value after every computation cycle. By default, initial value of reliability of each virtual node is 100 %. In this module reliability assessment algorithm (proposed) will be used to measure the reliability of virtual machines. However, if the reliability value of virtual machine will be less than the minimum reliability then the backward recovery will be performed by the system through fine- grained check pointing.

The proposed algorithm reliability assessment algorithm will depict the minimum and maximum reliability of nodes by identifying and replacing the failure nodes. The node status is checked for reliability analysis to fix the specific nodes for execution by eliminating the failure nodes. Step 3 shows node status, if the node status is passed it proceeds the process of IR= IR+(IR * RF).

IIOAB JOURNAL
ISSN: 0976-3104

Reliability Assessment Algorithm

Input :Initially reliability ← IR=1, n :=1 from configuration RF,
Step1: MR ←maxReliability
Step2: Mir←minReliability
Step3: Ns ← node status
Step4: If Ns = pass then
Step5: IR= IR+(IR * RF)
Step6: If n > 1 then
Step7: n=n-1
Step8: End if
Step9: Else if Ns = Fail then
Step10: IR =IR –(IR*RF*n)
Step11: n=n+1
Step12: End else if
Step13: If IR >= MR then
Step14: IR = MR
Step15: End if
Step16: C ← count=1;
Step17: While c<=4
Step18: C+1
Step19: If IR < Mir then
Step20: Ns = dead
Step 21: call_proc: remove_this_node
Step 22: call_proc: add_new_node
End if
End while

Virtual machine 0 to 9 is given in [Table 1]. The total time and reliability time is checked for each virtual machine to identify whether the machine performing its task with expected reliability time. In the table it is clearly shown that the 7, 8 and 9 failed to perform the task when its reliability time increases.

Table 1: Reliability time assessment

VMID	Total time	Reliability Time	Result
0	852	0.99734	Pass
1	114	1.994	Pass
2	366	2.992	Pass
3	954	3.9895	Pass
4	776	4.9869	Pass
5	407	5.984	Pass
6	747	6.981	Pass
7	562	7.979	Fail
8	134	8.976	Fail
9	137	9.9739	Fail

Fine-grained check pointing

The proposed fine grained Check pointing algorithm will work on the assumption that the length of the check pointing interval is not fixed and is calculated at the time of present checkpoint. This calculation is based on the failure history of VMs. If the failure rate is less, then the interval between checkpoints is reduced and if the node is having poor failure history, the interval between checkpoints is increases in the opposite scenario.

Fine –grained check point algorithm
Input: R←Reliability time
VM ←Virtual machine
T ←thresholds time
NC ← number of count=4
N ← Number count =1
B_{TF}←False
B_{Ts}←True
Step 1: While (!Task execution completed) do
Step 2: If (R>=T)then
Step 3: Estimation();
Step 4: END IF
Step 5: if(R <= T)

IIOAB JOURNAL
ISSN: 0976-3104

Step 6: R++;
Step 6: N++;
Step 7: If(N<=NC)
Step 8:then = true
Step 9: update the VM failure information.
Step 10: Create New VM
Step 10: end if
Step 11: end while

The proposed algorithm is implemented for checking and enhancing the reliability time of VM for completing the particular task efficiently. If the execution of the task halts suddenly, the reliability is checked by relating it with threshold values. If the reliability is greater than or equal to threshold time then the task will be finished within the schedule time without any interruption. If R<= T, the reliability and node count 4 is increased. If the N<=NC thus far, then BTF =true.

Then the VM failure information is updated.

Implementation

In the proposed framework Cloudsim is used to generate and evaluate the results [10]. The classes in the cloudsim package permit the development of algorithms based on fault tolerance which in turn can supervise virtual nodes towards the identification of failures and then later resolves them. This provision can deploy both fine grained check pointing and replication mechanisms. This simulator offers the capability to calculate availability, throughput, time overhead and monetary cost overhead. The timing of result produced by each virtual machine will be monitored with Watch Dog Timer Software. The performance measure was evaluated to prove the efficiency of proposed adaptive fault tolerance mechanism. The performance of the framework was evaluated in terms of execution time. The reliability assessment, decision making and the Fine grained check point algorithm proved their efficiency on the basis of outcome received.

Implementation of the proposed algorithms are shown as follows:

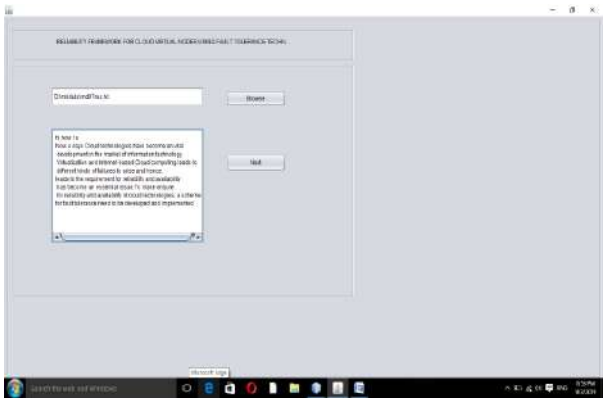

Fig. 2: input buffer

In [Fig. 2] reliability framework input has been taken from input buffer, after that virtual machines will be created and input will be given to the virtual machines.

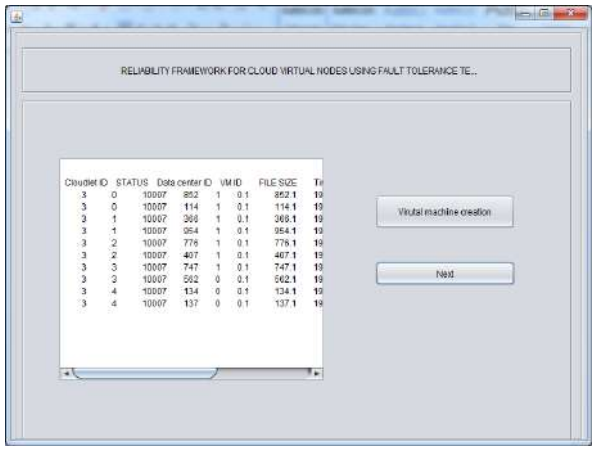

Fig. 3: VM creation

In [Fig. 3] ten virtual machines are created, after creation of virtual machines results will be checked whether it is correct or not.

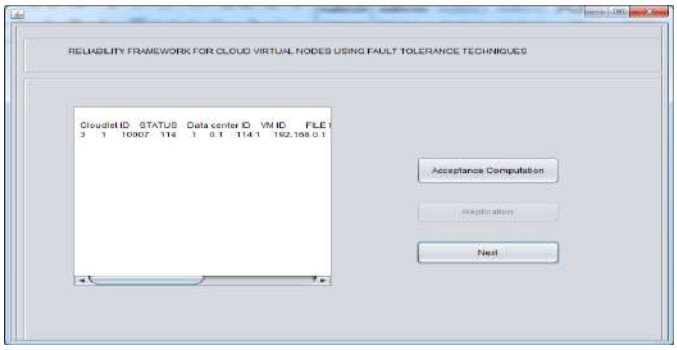

Fig. 4: Acceptance computation

In [Fig. 4], the total time and reliability time is checked for each virtual machine to identify whether the machine performing its task with expected reliability time. During the acceptance computation it is mentioned that the 7, 8 and 9 failed to perform the task when its reliability time increases. To increase their reliability fine-grained check pointing is used.

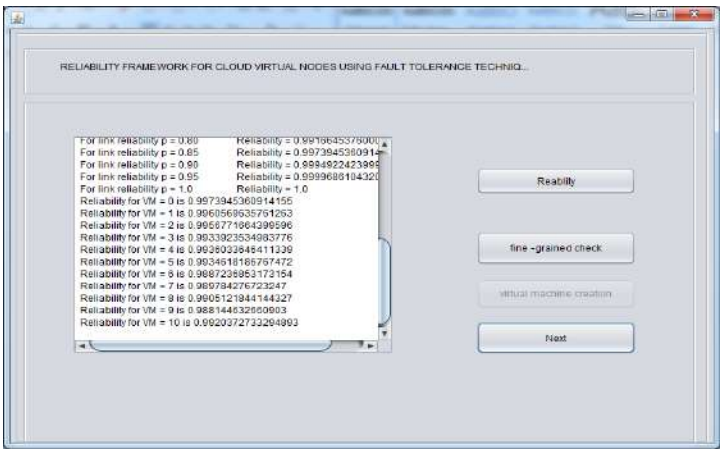

Fig. 5: Reliability framework using fault tolerance technique

In the above framework [Fig. 5], fine-grained check pointing is used to increase the reliability.

CONCLUSION

As we know that failures are inevitable in the cloud environment. To solve this problem, in this paper an adaptive fault tolerance method is proposed to provide reliability of cloud nodes. To handle the major challenge of tolerating fault related complexity, a study was done over cloud computing and introduced fault reduction scheme. In pre-emptive cases, the requests from users were influenced by this scheme for scheduling VMs. Also increment in using adaptive fault tolerance techniques provides reliability to the node outline. The reliability can also be increased by using fine-grained check pointing method which lowers node failures. The response time and availability is taken into account for performance enhancement. Backward recovery technique is be implemented to make the system fault tolerant. The adaptive nature of the proposed method specifies that the cloud's performance will be appreciably enhanced. The response time and availability is taken into account for performance evaluation. The reliability can also be increased by using replication method which can be implemented in upcoming research papers.

CONFLICT OF INTEREST
There is no conflict of interest

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] Dhingra M, Gupta G. [2017] Comparative analysis of fault tolerance models and their challenges in cloud computing. International Journal of Engineering and Technology, 6(2):36–40.

[2] Dhingra M, Gupta G. [2019] Architectural Framework for Cloud Reliability Model using Fault Tolerance Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(11):515-519.

[3] Mohammed A. [2016] Adaptive Framework for Reliable Cloud Computing Environment. IEEE Access.

[4] Limam S, Belalem G. [2014] A migration approach for fault tolerance in cloud computing. Int. J. Grid High Perform. Comput., 6(2):24–37.

[5] Cao J, Simonin M, Cooperman G, Morin C. [2015] Check pointing as a service in heterogeneous cloud environments Proc 15th IEEE/ACM Int. Symp Cluster, Cloud Grid Comput., Shenzhen, China, 61–70.

[6] Bobroff N, Kochut A, Beaty K. [2007] Dynamic Placement of Virtual Machines for Managing SLA Violations. Proc. of the 10th IFIP/IEEE International Symposium on Integrated Network Management, IM '07.

[7] Verrna A, Ahuja P, Neogi A. [2008] pMapper: Power and migration cost aware application placement in virtualized systems. Proc. of the 9th ACM/IFIP/USENIX International Middleware Conference.

[8] Sheheryar M, Fabrice H. [2011] Adaptive Fault Tolerance in Real Time Cloud Computing. IEEE World Congress on Services, 10:280-287.

[9] Meshram A, Sambare AS, Zade SD. [2013] Fault Tolerance Model for Reliable Cloud Computing. International Journal on Recent and Innovation Trends in Computing and Communication, 1(7):600-603.

[10] Damodhar M, Poojitha S. [2017] An Adaptive Fault Reduction Scheme to Provide Reliable Cloud Computing Environment. IOSR Journal of Computer Engineering (IOSR-JCE), 19(4): 64-73.