ISSN No.: 2393-963X

Journal of Advances in Computational Sciences and Information Technology

Volume No. 12

Issue No. 3

September - December 2024

ENRICHED PUBLICATIONS PVT.LTD

JE - 18,Gupta Colony, Khirki Extn,
Malviya Nagar, New Delhi - 110017.
E- Mail: info@enrichedpublication.com
Phone :- +91-8877340707

ISSN No.: 2393-963X

Journal of Advances in Computational Sciences and Information Technology

Aims and Scope

Journal of Advances in Computational Sciences and Information Technology research journal, which publishes top-level work from all areas of Computational Sciences and Technology. Areas and subareas of interest include (but are not limited to) Programming Languages; Software Development; Graphics for Science and Engineering; Solid, Surface and Wireframe Modelling; Animation; Data Management and Display; Image Processing; Flight Simulation; VLSI Design; Process Simulation; Neural Networks and their Applications; Fuzzy Systems Theory and Applications; Fault-Tolerant Systems; Visual Interactive Modelling; Supercomputing; Optical Computing; Soft Computing; Data Structures and Network Algorithms; Genetic Algorithms and Evolutional Systems; Very Large Scale Scientific Computing; Molecular Modelling; Scientific Computing in Emerging Critical Technologies; Computational Learning and Cognition; Computational Methods in Geosciences-Oceanographic and Atmospheric Systems; Computational Medicine; Artificial Intelligence; Cybernetics; Computer Security Issues; Evolutionary and Innovative Computing and more...

ISSN No.: 2393-963X

Journal of Advances in Computational Sciences and Information Technology

ISSN No.: 2393-963X

Journal of Advances in Computational Sciences and Information Technology

(Volume No. 12, Issue No. 3, September - December 2024)

Contents

Study the Role & Applications of Red Tacton - Technology for Communication Through Human Body

Harisha[1], Dr. Yashpal Singh[2]
Department of Electronics and Communication Engineering
[1,2]OPJS University, Churu (Rajasthan) – India

A B S T R A C T

NTT's (Nippon Telegraph and Telephone Corporation) labs, Japan has reported new technology called "Red Tacton, it is a Perfect networking technology that transfers data to devices through people's bodies with high speed. Communication is conceivable utilizing anyone surfaces, for example, the hands, fingers, arms, feet, face, legs or toes. It works through shoes and garments too. At the point when the physical contacts get isolated, the communication is isolated. Utilizing RedTacton, the electric fields created by the human body function as a mode for transmitting the data. The chips which will install in different devices contain a transmitter and recipient worked to send and acknowledge data in digital format. The chips can take any sort of document, for example, mp3 music record, or mail and convert in to the digital arrangement that can be gone and read through a human being as electric field. The chip in beneficiary devices peruses these modest changes and converts the document over into its unique arrangement. NTT grew very sensitive Photonic electric field sensor for recognizing minute electric field discharged on the surface of the human body. In this article we have studied the role & applications of red tacton - technology for communication through human body.

1. NTRODUCTION

In practical usage, a RedTacton device must act both as a transmitter and a beneficiary (i.e., like a transceiver). The interface sends a data signal to the two Data sense circuit and transmitter circuit. On the off chance that the data is available, the data sense circuit senses the data and a control single are created and send a control signal to the transmitter consequently; the transmitter circuit is triggered. Transmitter circuit creates the electrical signals that are to be accepted into the human body and is sent to the transmitting/getting cathode. Since, the terminal capacitively coupled to the human body, transfer of electrical signal happens.

At whatever point a connection is built up by touching, the anode present at the recipient side distinguishes the signal and feeds it to the electro-optic sensor circuit. Without the control signal from the data sense circuit, the yield is nourished into locator circuit this recognizes transmitting and getting modes to enable two- way communication. At long last, the yield of the finder circuit is sent to the interface.

Multiple REDTACTON transceivers can be utilized on the double since it claims the CSMS/CD (Career Sense Multiple Access with Collision Detection) conventions that send simply in the wake of checking the medium, to ensure that there is no data to be gotten to keep away from bundle collisions[49].

2. SECURITY ENHANCEMENTS IN RED TACTON – HUMAN AREA NETWORKING TECHNOLOGY

Amid the most recent decade, well being mindfulness and well being control have expanded among all gatherings of human creatures and the total need to watch biomedical information every day, medicinal services the executives and sensor investigation has expanded. There is an interest to refresh the knowledge, productivity, execution and security of personal and sensitive information of biomedical data sensors and to interface them with the Internet of Things (IOT) framework for the steady working of the network around the body significant. It is additionally important to modify the administration of biomedical information through the insurance and security of protection. This thusly makes the requirement for communication at short proximity whichstarts in the human body, rather than long-extend radio communications or different means.

The human body communication technology (HBC), which utilizes the human body as a methods for signal transmission, is one of the promising body network technologies (BAN) utilized for communication all through the human body. This investigation centers around HBC technology and concentrates the transmission attributes and dissemination of the electric field around the human body. Moreover, the transmission of heartbeat data utilizing HBC technology is likewise examined. It is totally not the same as wireless and infrared technologies, since it utilizes the little electric field discharged on the surface of the human body.

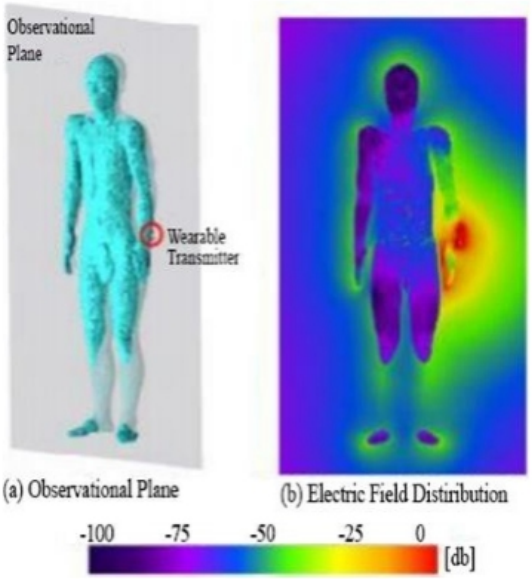


Figure 1: Electric Field Distribution with Transmitter Mounted

The electric field sensor (transistor or photonic electric field sensor) recognizes electric field that achieves the Red Tactonreceiver. The recipient senses the adjustments in the feeble electric field on the surface of the body caused by the transmitter. Red Tacton depends upon the rule that the optical properties of an electro-optic crystal can differ as indicated by the progressions of a feeble electric field.

The normally happening electric field actuated on the surface of the human body disseminates into the earth. Consequently, this electric field is uncommonly blackout and unstable. The photonic electric field sensor created by NTT enables feeble electric fields to be measured by recognizing changes in the optical properties of an electro-optic crystal with a laser beam.

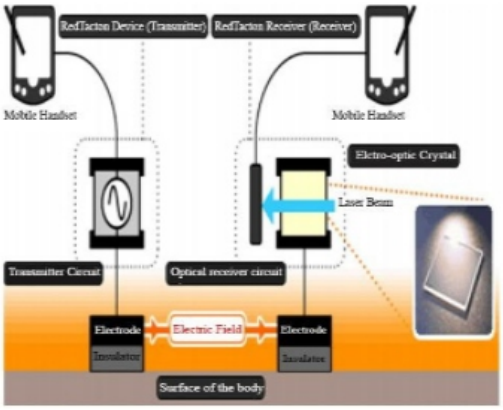


Figure 2: Working principle of Red Tacton

3. NEW TECHNOLOGY TO USE HUMAN BODY AS DIGITAL TRANSMISSION BOTH

Nippon Telegraph and Telephone Corporation (NTT) is seeking after research and advancement of an inventive Human Area Networking technology called RedTacton (*1) that securely transforms the surface of the human body into a data transmission path at speeds up to 10 Mbps between any two points on the body. Utilizing a novel electro-optic sensor (*2), NTT has officially built up a little PCMCIA card-sized model RedTacton transceiver. RedTacton enables the primary practical Human Area Network between body- centered electronic devices and PCs or other network devices inserted in the earth using a new generation of UI based on normal human activities, for example, touching, holding, sitting, strolling, or venturing on the spot.

RedTacton can be utilized for the instinctive task of PC based frameworks in day by day life, impermanent coordinated private networks based on personal handshaking, device personalization, security, and a large group of different applications based on new standards of conduct enabled by RedTacton. NTT is focused on moving RedTacton out of the research facility and into business generation as fast as conceivable by sorting out joint field preliminaries with accomplices outside the organization, under NTT's far-reaching maker program. Human society is entering a time of pervasive computing when networks areconsistently interconnected, and the information is constantly accessible at our fingertips.

The practical execution of pervasive administrations requires three dimensions of availability: Wide Area Networks (WAN), regularly by means of the Internet, to remotely interface a wide range of separates and terminals; Local Area Networks (LAN), ordinarily through Ethernet or WiFi availability among all the information and communication apparatuses in workplaces and homes; and Human Area

among all the information and communication apparatuses in workplaces and homes; and Human Area Networks (HAN) for network to personal information, media and communication machines inside the lot littler circle of normal every day exercises the last one meter.

NTT's RedTacton is a leap forward technology that, out of the blue, enables dependable high- speed HAN. Before, Bluetooth, infrared communications (IrDA), radio frequency ID frameworks (RFID), and different technologies have been proposed to settle the "last meter" network problem. In any case, they each have different central specialized impediments that oblige their use, for example, the abrupt tumble off in transmission speed in multi-client situations creating network congestion.

TECHNICAL POINT

RedTacton takes a different technical approach. Instead of relying on electromagnetic waves or light waves to carry data, RedTacton uses weak electric fields on the surface of the body as a transmission medium. A RedTacton transmitter couples with extremely weak electric fields on the surface of the body. The weak electric fields pass through the body to a RedTacton receiver, where the weak electric fields affect the optical properties of an electro-optic crystal. The extent to which the optical properties are changed is detected by laser light which is then converted to an electrical signal by a detector circuit. The three major functional features of RedTacton are highlighted below.

1. A communications path can be created with a simple touch, automatically initiating the flow of data between a body-centric electronic device and a computer that is embedded in the environment. For example, two people equipped with RedTacton devices could exchange data just by shaking hands. A wide range of natural human actions -- grasping, sitting down, walking, or standing in a particular place -- can be used to trigger RedTacton to start a networked process.
2. Using a RedTacton electro-optic sensor, two-way communication is supported between any two points on the body at a throughput of up to 10 Mbps. Communication is not just confined to the surface of the body but can travel through the user's clothing to a RedTacton device in a pocket or through shoes to communicate with a RedTacton device embedded in the floor.
3. RedTacton can utilize a wide range of materials as a transmission medium, if the material is conductive and dielectric, which includes water and other liquids, various metals, certain plastics, glass, etc. Using ordinary structures such as tables and walls that are familiar and readily available, one could easily construct a seamless communication environment at very low cost using RedTacton.

RedTacton adopts an alternate specialized strategy. Rather than relying upon electromagnetic waves or light waves to convey data. RedTacton utilizing feeble electric fields on the surface of the body as a transmission medium as appeared in figure 3.

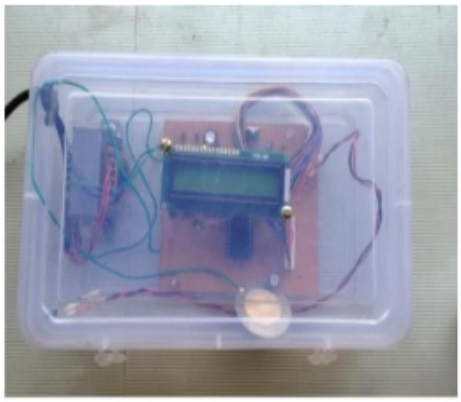


Figure 3: Transmitter

1. The RedTacton transmitter induces a weak electric field on the surface of the body.
2. The RedTacton receiver senses changes in the weak electric field on the surface of the body caused by the transmitter.
3. RedTacton relies upon the principle that the optical properties of an electro-optic crystal can vary according to the changes of a weak electric field.
4. RedTacton detects changes in the optical properties of an electro-optic crystal using a laser and converts the result to an electrical signal in a optical receiver circuit.

Applications - Minimization of human mistake, advertising applications, customization of vehicles, modern checking.

- Security Applications - Automatic user authentication and log-in with just a touch.
- ID and privileges are recorded in a mobile RedTacton device.
- Corresponding RedTacton receivers are installed at security check points.
- The system can provide authentication and record who touched the device
- User Verification Management: Carrying a mobile RedTacton-capable device in one's pocket, ID is verified, and the door unlocked when the user holds the doorknob normally
- Secure lock administration is possible by combining personal verification tools such as fingerprint ID or other biometric in the mobile terminal.

4. REDTACTON NEAR-BODY ELECTRIC- FIELD COMMUNICATIONS TECHNOLOGY AND ITS APPLICATIONS

The near-body electric-field communication system described in this Special Feature consists of a transmitter that emits an alternating current (AC) electric-field signal modulated by input data from an electrode and a receiver that uses another electrode to read the weak AC electric field on the body induced by the signal and demodulates it to recover the data from it (Figure 4).

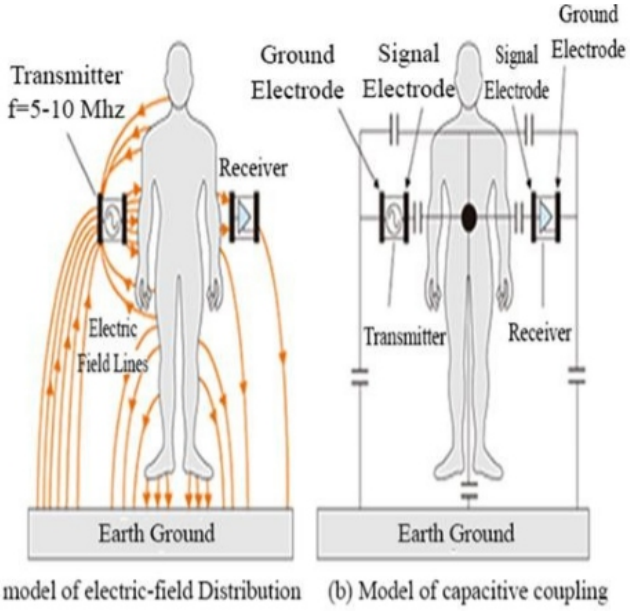


Figure 4: Electric-field and Capacitive-Coupling Models for Near-Body Electric-Field Communications

5. CONCLUSION

We have concluded that the transmitter and recipient are capacitive coupled to the human body through level cathodes that are proportional to the antennas in ordinary wireless frameworks. The close body electric field is adjusted and can be perused and demodulated through these couplings. A component of the technology is that, on the grounds that the signal is passed to and from the body's surface through this capacitive coupling, it tends to be transmitted among transmitter and recipient by means of the body's surface, regardless of whether one of them is in the client's pocket, and the other is under the cover on the floor, for instance. If this technology is connected to a section control framework or ticket door, it is considerably more helpful than customary Contactless cards since you don't have to expel the access card from your pocket.

6. REFERENCES

[1]. Mario CifrekZeljkaLucev, Igor Krois. Application of wireless intra body communication system to muscle fatigue monitoring. ETRI Journal, 2010.

[2]. M. Shinagwa, M. Fukumoto, K. Ochiai, and H. Kyuragi, "A nearfield- sensing transceiver forintra-body communication based on the electro- optic effect," IEEE Trans. Instrum. Meas., vol.53, no.6, pp.1533–1538, 2004.

[3]. S.-J. Song, N. Cho, S. Kim, J. Yoo, and H.-J. Yoo, "A 2Mb/s wideband pulse transceiver with direct-coupled interface for human body communications," Digest of Tech. Papers IEEE Int. Solid- State Circuits Conf. (ISSCC), pp.558– 559, San Francisco, USA, Feb. 2006.

[4]. N. Cho, J. Yoo, S.-J. Song, S.-J. Lee, S. Jeon, and H.-J. Yoo, "The human body characteristics as a signal transmission medium for intra body communication," IEEE Trans. Microw. Theory Tech., vol.55, no.5, pp.1080–1086, 2007.

[5]. A. Fazzi, S. Ouzounov, and J. Hombergat, "A 2.75mW wideband correlation-based transceiver for body-coupled communication," Digest of Tech. Papers IEEE Int. Solid-State Circuits Conf. (ISSCC), pp.204–205, San Francisco, USA, Feb. 2009.

[6]. Y. Kado, "Human-area networking technology as a universal interface," 2009 Symposium on VLSI Circuits Digest of Technical Papers, pp.102–105, Kyoto, Japan, June 2009.

Study of Role of Machine Learning Techniques in Radio Networks

Abhishek Kumar Gupta[1], Dr. Yashpal Singh[2]
Department of Computer Science and Engineering
[1,2]OPJS University, Churu (Rajasthan) – India

ABSTRACT

Cognitive radio offers the promise of intelligent radios that can gain from and adapt to their environment. To date, most cognitive radio research has concentrated on policy-based radios that are hard-coded with a rundown of principles on how the radio ought to act in specific situations. Some work has been done on radios with learning engines custom fitted for unmistakable applications. This article depicts a solid model for generic cognitive radio to use a learning engine. The objective is to fuse the consequences of the learning engine into a predicate math-based reasoning engine with the goal that radios can recollect exercises learned previously and acted rapidly.

1. INTRODUCTION

Decision-making and feature classification. Decision-making is in charge of deciding policies and decision rules for CRs while feature classification grants are distinguishing and ordering diverse perception models. The learning calculations experienced are sorted as either regulated or unsupervised calculations.

We portray in detail a few testing learning issues that emerge in cognitive radio networks (CRNs), specifically in non-Markovian environments and decentralized networks, and present conceivable arrangement strategies to address them. We examine likenesses and contrasts among the presented calculations and recognize the conditions under which every one of the techniques might be connected.

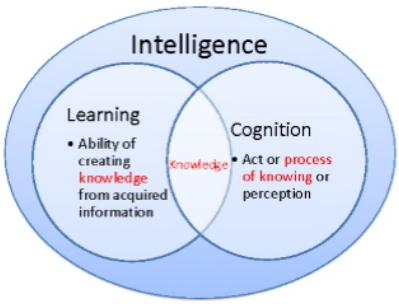


Figure 1: Artificial Intelligence Communication System

2. MACHINE LEARNING TECHNIQUES FOR COOPERATIVE SPECTRUM SENSING IN COGNITIVE RADIO NETWORKS

The idea of Cognitive Radio (CR) for designing wireless communications systems has emerged since a decade ago to moderate the scarcity problem of restricted radio spectrum by enhancing the use of the

spectrum. The CR alludes to an intelligent wireless communications device, which detects its operational electromagnetic environment and can dynamically and independently modify its radio operating parameters. In this specific situation, Opportunistic Spectrum Access (OSA) is a key idea, which enables a CR device to opportunistically access the recurrence band dispensed to a primary client (PU) when the PU transmission is detected to be inactive[3]. For OSA, the CR devices need to detect the radio spectrum authorized to the PUs by utilizing its restricted resources (e.g., vitality and computational power), and in this manner use the accessible spectrum chances to amplify its execution goals. Consequently, proficient spectrum detecting is essential for OSA. Because of the quick advancement of the current Internet and Mobile Communication industry, the mobile movement stack has encountered a touchy development amid the most recent couple of years too. In like manner, the systems of mobile network operators (MNOs) have been planned and enhanced with greater unpredictability of infrastructure, higher decent variety of related devices and resources, and more dynamical arrangements of networks, and they are advancing towards the promising future mobile networking paradigm, the heterogeneous networks (HetNets).

AI techniques incorporate multi-disciplinary techniques from machine learning, bio-enlivened calculations, and fluffy neural network et cetera, and they have been broadly examined and connected to enhance PC systems and networks in different situations and confounded environments. It has been demonstrated that AI techniques can accomplish exceptional execution, as a large portion of them are roused from nature discoveries or spurred by the mindsets of people. They have generally brought down multifaceted nature empowered by recursive input based learning and neighborhood cooperation's, and subsequently speedier speed of discovering imperfect arrangements contrasted with ordinary techniques.

3. MACHINE LEARNING TECHNIQUES

Machine learning (ML) is advanced from the study of example acknowledgment and computational learning hypothesis in AI regions. It viably takes in the method for human brains, investigates the development and study of calculations, and settles on information driven expectations or decisions. Among numerous ML techniques, the reinforcement learning (RL) is one specific learning calculation roused by behaviorist brain science, worried about how programming specialists should take activities in an environment to expand some idea of combined prizes.

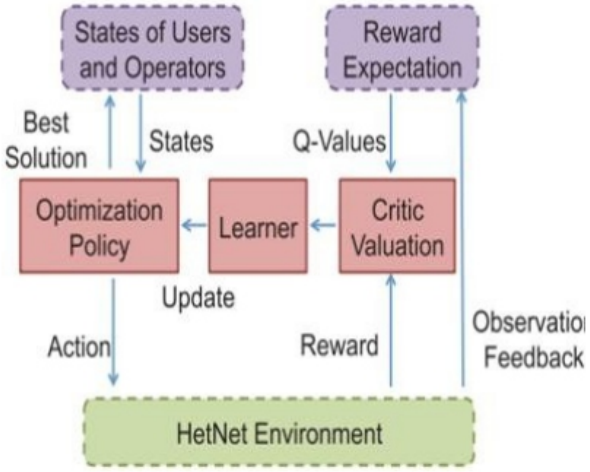


Figure 1.2: Illustration of Learning Based Optimization in Hetnets

4. AI TECHNIQUES IN NETWORKING

In the networking space, this learning would incorporate models of material science and signal spread, requirements on the system, investigation of collaborations, and dependable guidelines (e.g., about how to design the system). A formal cosmology may enable a cognitive system to reason about how and when capacities are exchangeable, e.g., perceiving that both of two measurements for figuring Quality of Information might be utilized and that a metric for Quality of Service might be a proper substitution under a few conditions. Semantics and representations are critical contemplations for cognitive networks. A few researchers have created information bases and heuristic standards to streamline the network. Arranging and planning techniques are proper for decision- making circumstances, where errands should be composed and facilitated to meet execution destinations, under asset imperatives. In powerful environments, the arrangement should be checked because expectations about execution may have been off base or the conditions have changed with the end goal that already chose activities are never again suitable.

They likewise talked about their qualities, shortcomings, and the challenges in applying these techniques in CR undertakings. In [14], the creators thought about diversion hypothesis, fortification learning, and thinking methodologies, for example, Bayesian networks, fluffy rationale, and case-based thinking. As opposed to the writing, we present a far-reaching review considering all the learning techniques that were utilized in cognitive networks. The study is composed dependent on various artificial intelligence approaches including the accompanying:

1. Fuzzy logic,
2. Genetic algorithms,
3. Neural networks,

4. Game theory,
5. Reinforcement learning,
6. Support vector machine,
7. Case-based reasoning,
8. Decision tree,
9. Entropy,
10. Bayesian,
11. Markov model,
12. Multi-agent systems, and
13. Artificial bee colony algorithm.

5. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence aims at making machines perform tasks in a manner similar to an expert. The intelligent machine will perceive its environment and take actions to maximize its own utility. The central problems in artificial intelligence include deduction, reasoning, problem solving, knowledge representation, and learning:

1. Sensing the radio frequency (RF) parameters such as channel quality,
2. Observing the environment and analyzing its feedback such as ACK responses,
3. Learning,
4. Keeping the decisions and observations for updating the model and obtaining better accuracy in future decision- making, and finally
5. Deciding on issues of resource management and adjusting the transmission errors accordingly.

The fuzzy set theory was proposed by Lotfi A. Zadeh in 1965 to solve and model uncertainty, ambiguity, imprecision's, and vagueness using mathematical and empirical models. The variables in fuzzy logic are not limited to only two values (True or False) as it is defined in classical and crisp sets. A fuzzy element has a degree of membership or compatibility with the set and its negation. Fuzzy logic provides the system with

1. Approximate reasoning by taking fuzzy variables as an input and producing a decision by using sets of if-then rules,
2. Decision-making capability under uncertainty by predicting consequences,
3. Learning from old experience, and
4. Generalization to adapt to the new situations.

Table 1: Learning Techniques Evaluation: Strengths, Limitations, And Challenges

Learning technique	Spectrum sensing (SS)	Decision- making	Strengths	Limitations and challenges
			Adaptation ability to minor changes	Require training data labels
Neural networks	×	×	Construction using few examples, thus reducing the complexity	Poor generalization Over fitting
Support			Generalization ability	Requires training data labels
vector	×	×	Robustness against noise and outliers	and previous knowledge of the system
machine				Complex with large problems
			Multi-objective optimization	Require prior knowledge of the system
Genetic algorithms		×	Dynamically configure the CR	Suitable fitness function
			based on environment changes	High complexity with large problems
Game theory	Related to the capabilities of the	×	Reduces the complexity of adaptation	Requires prior knowledge of the system
	spectrum-sensing technique used		Solutions for multi- agent systems	and labeled training data
Reinforcement	×	×	Learning autonomously using feedback	Needs learning phase of the policies
learning			Self-adaptation progressively in real time	
Fuzzy logic	Related to the capabilities of the	×	Simplicity, decisions are	Needs rules derivation
	spectrum-sensing technique used		directly inferred from rules	Accuracy is based on these rules
Entropy approach	×	×	Statistical model	Requires prior knowledge of the system
	Related to the capabilities of the		Simplicity	Requires prior knowledge of the system
Decision tree	spectrum-sensing technique used	×	Decision using tree branches	May suffer over fitting
				Requires labeled training data
Artificial		×	Parallel search for solutions	Requires prior knowledge of the system
bee colony				Requires a fitness function
Bayesian	×	×	Probabilistic models	Requires prior knowledge of the system
				May face computational complexity
Markov model	×	×	Statistical models	Requires prior knowledge of the system

6. MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE FOR MANAGING COMPLEXITY

The ML and AI are two powerful tools that are developing as solutions for managing a lot of data, particularly to make predictions and giving proposals based on the data sets. They are; be that as it may, all the time seemed, by all accounts, to be utilized reciprocally despite a few parallels. ML is sometimes raised as a subspace of AI based around the idea that we can give the machines a chance to learn for themselves by giving them access to a lot of data. Then again, AI is the all-encompassing and more extensive impression of machines getting to be fit for completing tasks in an astute way. Contrasted with the summed up AI (a summed up AI system, in theory, can deal with any task), connected AI is progressively appropriate for cutting edge communication systems as the connected AI system can be conceived to adroitly controlling and enhancing the wireless networks. Dissimilar to ML models, AI models connect the world, familiarize to the progressions and rebuild themselves.

- **Subscriber Mobility Pattern (SMP):** In order to guarantee the QoS requirements and to efficiently maintain resources utilization, traffic offloading and routing, knowing the mobility information of a user in advance is very crucial. Human travel pattern analyses reveal that people travel along specific paths with reasonably high predictability. The trajectory of a mobile user can be predicted based on user's present location, the movement direction and the aggregate history of SMP.

- **Radio Environment Map (REM):** The MNOs can better plan, build, control and optimize their networks conforming to the spatiotemporal radio atmosphere, through prediction of radio signal attenuation. Many schemes have been developed that give the MNOs the means to predict the distribution of radio signal attenuation at different operating frequencies and in many different radio environments. The radio map along with the mobile user's predicted trajectory facilitates the prediction of average channel gains.

- Traffic Profile (TP): In order to attain as well as predict the network's congestion status, tempo-spatial traffic load variation needs to be known, i.e., the knowledge of temporal traffic trace, BS spatial deployment and BSs' operating characteristics (transmission power, height, etc.) are very important. The authors report that the network's traffic load dynamics demonstrates periodical characteristics over days and hours, thus implying high predictability of the traffic load.

7. CONCLUSION

We have concluded that the intelligence of cognitive radio technology and machine learning offers the potential to learn and adjust to the wireless environments. As the utilization of machine learning

techniques in wireless communications is normally joined with cognitive radio technology, we have concentrated on both cognitive radio technology and machine learning to give a complete diagram of their jobs and relationship in accomplishing intelligent wireless communications. We have thought about range productivity and vitality proficiency, the two of which are vital qualities of intelligent wireless communications.

REFERENCES

[1]. B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," IEEE Circuits Syst. Mag., vol. 9, no. 2, pp. 27–45, Jun. 2009.

[2]. Laursen, K., Foss, N. (2003). New human resource management practices, complementarities and the impact on innovation performance, Cambridge Journal of Economics, 27, p. 243 – 63.

[3]. Laursen, K., Foss, N. (2003). New human resource management practices, complementarities and the impact on innovation performance, Cambridge Journal of Economics, 27, p. 243 – 63.

[4]. Lazear, E.P. (2000), "The power of incentives", The American Economic Review, Vol. 90 No. 2, p. 410-15

[5]. McConville, T. (2006). Devolving HRM responsibilities, middle managers and role dissonance. Personnel Review 35, no. 6: 637–53.

[6]. Karatepe, O., Arasli, H., & Khan, A. (2007). The Impact of Self-Efficacy on Job Outcomes of Hotel Employees. International Journal of Hospitality & Tourism Service, 8(4), 23–46. doi:10.1300/J149v08 n04_02.

[7]. Niu, H.-J. (2010). Investigating the effects of self-efficacy on foodservice industry employees" career commitment. International Journal of Hospitality Management, 29(4), 743– 750. doi:10.1016/j.ijh m.2010.03.006.

[8]. Walumbwa, F. O., &Hartnell, C. A. (2011). Understanding transformational leadershipemployee performance links: The role of relational identification and self-efficacy. Journal of Occupational and Organizational Psychology, 84(1), 153–172. doi:10.1348/096317910X485818.

[9]. Sheehan, C. (2005). A model for HRM strategic integration. Personnel Review, 34(2), 192 – 209. doi:10.11 08/00483480510579420.

Review and Simulation of Different Sampling Methods

Christian Hoops[1], Rahul Pathare[2]
[1]Ipsos Public Affairs, Sachsenstraße 6, 20097 Hamburg (Germany)
[2]Narsee Monjee Institute of Management Studies, University of Mumbai
V.L. Mehta Road, Vile Parle, Mumbai – 400 056, Maharashtra, India

A B S T R A C T

Obtaining information about hard-to-reach populations is a major challenge in the market research field. In our case of analyzing German minorities in Denmark, only a small fraction of the total population belongs to the target population. Therefore, selecting minorities by generating telephone numbers at random would result in very high costs. Alternative sampling methods have to be used, but there are no practices to identify the best approach. This article tries to fill this gap and creates a comparison of snowball sampling (SS), random digit dialing (RDD), gravity sampling (GS) and facility-based sampling (FBS). Sample data has been extracted by a previous survey (Hoops, Schnapp and Schaefer-Rolffs 2013) and a further model extended by randomly generating and simulating all four sampling methods using bootstrapping procedures. This enabled us to estimate the cardinality of the sample space, the bias and the variance of the inclusion probabilities in the sample for each method. Only GS and RDD create samples which are asymptotically unbiased. The combination of gravity and complete as well as non-overlapping citizens registers produces the highest cardinality of the sample space. But in contrast to RDD, citizens registry-office methods allow no household samples. So gravity analyses help to identify regions with a high prevalence of the target population to create samples with roughly varying inclusion probabilities. Our simulations indicate that gravity sampling methods using official databases produce very high quality samples. For cost reasons this method should be tested in practice to conduct surveys with hard-to-reach populations.

Keywords: Bootstrapping, gravity sampling, snowball sampling, surveys

1. NTRODUCTION

Surveys of hard-to-reach populations have always been a great challenge to scientists. Traditional methods of sampling these populations are ineffective and have many bias problems (Agadjanian and Zotova 2012). Researchers who want to create samples by using established methods like RDD, Marpsat and Razafindratsima (2010) state that the small percentage of the target population will increase the cost of investigation.

Researchers require many respondents for representative statements, which is quite difficult to collect from hard-to- reach populations and, on top of that, the risk of missing important data is slightly higher than in standard research due to marginal group reasons. In addition, respondents might not be able to recommend accurate target samples for further study due to a lack of knowledge or non-willingness to suggest eligible people (Hendricks and Blanken 1992).

However, the term “hard-to-reach populations” has to be delimited from similar terms like “hidden populations” who do not want to reveal their identity, such as HIV-infected persons or prostitutes. It has been seen in many research papers that these hidden populations are not willing to reveal their identity because they might be stigmatized by society and do not wish to recall horrible past events (Faugier and Sargeant 1997; Hendricks and Blanken 1992; Marpsat and Razafindratsima 2010). In contrast, hard-to-reach populations (e.g. immigrants or national minorities) are only defined by a low prevalence. There are several methods of identifying hard-to-reach populations, such as time-location sampling, snowball sampling and facility-based sampling. However, these methods can only generate samples of unknown composition.

This article mainly focuses on national minorities – those who feel that they belong to Germany, but are living in Denmark. With the Bonn-Copenhagen Declarations of 1955 (see Jäckel 1959: 74ff.), it is specified to protect this small group and not to conduct the minority status in official surveys. However, political parties of minority people have strong influence during elections and subsequently in their parliament. Perceptions and views of minority people play an important role during elections, thus it is important to find these minority populations and collect information by interviewing them via appropriate sampling methods (Hoops, Schnapp and Schaefer-Rolffs 2013). However, such groups make up only a small percentage of the country‟s total population. This means that the prevalence which depends on the sampling method is very low.

There are many approaches available in this research field but, unfortunately, there is no systematic or an established methodology that can accurately suggest which technique to select and apply in the study of low prevalence. Moreover, there are limited researches available which compare survey methods in practice to identify differences between them. This paper tries to fill this research gap.

First, we will give a review of different sampling methods to conduct with hard-to-reach populations. After that we will simulate the method-specific bias, the cardinality of the sample space and the variance of the inclusion probabilities in the sample. The sample space consists of all people with target characteristics who could be included in the sample. From this it follows that the quality of the sample space can be measured as the closeness of the sample space to the target population. The cardinality of the space sample indicates this degree of consistency.

2. LITERATURE REVIEW

2.1 Random Digit Dialing

Telephone method sampling is done by randomly selecting numbers from the telephone directory and making calls to a person (Brunner and Brunner 1971) for an interview. The method was invented in the USA where out of 10 digits the first 3 digits are the area code, then 3 digits are the telephone switching center and 4 digits are the suffix. In RDD, after entering the first six digits in the system, the last four digits are selected randomly.

On the other hand, in Germany, the first 3 to 5 digits (e.g. 040 or 04542) are the regional area code, which is followed by some digits, out of which the last two digits are used to generate a random number. However, unfortunately, regional sampling is not possible for mobile phones.

Table 1: Example of random digit dialing

Area code	*Number block*	*Status*
4542	55832 00	Registered
4542	55832 **01**	Generated
4542	55832 02	Registered
4542	…	…
4542	55832 99	Registered

In random digit dialing, strictly speaking, all non-registered RDD (Frankel et al. 2007). They have further described the significance of weighting telephone surveys that include cellular phones as well as landline telephones.

It is important to apply an accurate sampling and response rate calculation method in survey analysis (Ezzati-Rice et al. 2000). The argument is that the eligibility of respondents is related with non-response of selected interviewees and, to support this statement, Ezzati-Rice et al. (2000) have discussed the advantage of combining the response and coverage rate to calculate the response rate. In RDD, the non-response bias can be calculated by the response rate of a survey. However, while calculating the response rate, the researcher should exclude all non-coverage respondents who are eligible but do not have a telephone or mobile at home (Glasser & Metzger 1972; Frankel et al. 2007).

Massey & O'Connor (1997) have recommended calcu-lating the response rate at every stage of the sample design and also considering new types of incentives for the interviewer and interviewee. They further encouraged the researchers to develop new methodology for a better response rate and to perform research on the effects of questionnaires on respondents.

2.2 Snowball Sampling

It is important to study subjects with sensitive issues, like HIV study (Faugier and Sargeant, 1997; Kendall, et al. 2008), drug addicts (Lopes, Rodrigues and Sichieri 1996), migrantsnumbers between the lowest and the highest will be determined (see Fuchs 1994). However, this approach is mostly not practical, especially when many generated numbers do not exist. Therefore the randomized last digit method is being performed, where the last digits are replaced by randomly generated digits. Since the shares from registered and generated numbers are not identical in each number block, this method produces strongly varying inclusion probabilities in the sample. Blocks with a large amount of registered numbers are more often included in the sample than blocks with few registered numbers. With appropriate corrections, these selection differences can be compensated.

According to Glasser & Metzger (1972), the telephone survey method (see Table 1) is less expensive and has a better rate of response, with the callback option or monitoring previous calls, as compared to personal contacts or mail survey. However, random digit dialing is limited to people who possess a landline or mobile phone and creates a non- response bias (Glasser and Metzger 1972). The article by Esslemont et al. (1992) compares demographic character- istics between listed and unlisted respondents and found out that, due to telephone directory sampling, differences exist between them.

Previously, usage of cellular sampling was excluded from the RDD sampling because of the lower number of mobile phones users, and for ethical and legal issues. However, in the last decade, mobile phone usage has increased tremendously, and as a result survey companies include mobile sampling in (McKenzie and Mistiaen 2009) etc. To obtain necessary prerequisite data for research and study of hard-to-reach or hidden population, snowball sampling is one of the most efficient method available (Lopes et al. 1996, Berg 1983). The snowball sampling approach suggests that a new respondent can be created by a referral series within a group of people, who know each other, which then forms multiple recommendation waves and develops interesting comparisons in the group.

Berg (1988) posits that if a sampling frame is unavailable to cover the population, respondents‟ knowledge and connections are necessary to start the referral chain process. The chain referral technique is a self-propelled phenomenon (Biernacki & Waldorf 1981), which, once it gets started, automatically works by itself. During the process the scientist should first search for the respondents as an initial referral chain then verify their eligibility according to the proposed research subject. Furthermore, he must encourage all initial respondents to become research assistants in the project and ask them to recommend the next person for the study. In addition, he should manage, control and limit the chain process in order to avoid excessive, unnecessary respondents, and then motivate them from

time to time to maintain the quality of the data (Biernacki & Waldorf 1981; Marpsat & Razafindratsima 2010).

A comparative survey investigated by McKenzie & Mistiaen (2009) to analyze data of Japanese-Brazilian families as well as a HIV study by Faugier & Sargeant (1997) have found out that snowball sampling has a wide scope to conduct qualitative sociological research and has an edge over other sampling methods when it comes to reaching hidden populations in society. It is a very efficient, economical method which produces results in less time (Atkinson and Flint 2001).

Even though the snowball technique is an effective sampling method for the study of rare population, it has some limitations. Due to social problems like stigmatization from society, hidden population subjects were reluctant to reveal their identity (Faugier and Sargeant 1997; Hendricks and Blanken 1992). Berg (1988) claims, that if respondents are not guaranteed the confidentiality of information about them, ethical issues can arise. Other than social and ethical problems, Biernacki and Waldorf (1981) argue that the snowball technique is an expensive methodology as the interviewer is responsible from the start to end of the research. During the research period the analyzer should be concerned about information confidentiality, for example some people from the same referral chain could meet on the street, in court or in prison, and any disclosure of information may cause a serious problem in the entire research study (Berg 1988). The survey method in an urban populated area could be affected because of sampling error by clustering, interviewers‟ bias, bad weather and imbalanced interviews of homebound, homeless and unemployed people (Miller et al. 1997).

Illenberger et al. (2008) found out that the overestimation bias problem occurred in the early stage of snowball sampling and the results are limited to mean degree and the clustering coefficient. In their research study they try to find out: How initial respondents affect the subsequent chain process as well as consequences of respondents‟ refusal to participate in the survey. Furthermore, they encourage other researchers to investigate bias-correction measures of “closeness and betweenness”.

The recent paper by Johnston and Sabin (2010) on hard-to- reach population shows that the respondent-driven sampling method by Heckathorn (1997) can be a very good alternative to reduce initial bias problems associated with snowball sampling. They argue that respondents of snowball sampling give unlimited references which create bias issues such as clustering, differential recruitment and accessibility to further samples. This can be eliminated by issuing coded coupons that can be redeemed only at a fixed location (interviewer‟s office) in a limited time. Respondents are allowed to refer only a limited number (2-3) of target people, which further reduces the over representation bias problem.

Moreover, Johnston and Sabin (2010) suggest that respondents should be rewarded with gifts or shopping coupons, rather than cash, to motivate them to participate in the study. The interviewer should estimate the price of the gift in such a way that it should not be overvalued or undervalued. Because, if the gift price is overvalued, then the respondent may sell it to another person and if it is undervalued, then the respondent will not show up for the interview. At least the respondent-driven technique allows to make asymptotically unbiased estimates (Salganik and Heckathorn 2004). Nevertheless, the sample space has a low cardinality due to target people with only a small or without social network.

2.3 Facility-Based Sampling

The success of hard-to-reach sampling depends on available information and knowledge about the target population. Snowball sampling, response-driven sampling, and time location sampling are a few examples of how researchers are able to approach rare population (Sangngam and Suwattee 2010). Researchers can also use the facility-based survey method to find hard-to-reach people, in which the interviewer selects a specific location (e.g. hospital) where he gets information about the target population (e.g. drug addicts) and he can also use different facilities of the location and then the researcher can conduct survey interviews with doctors, nurses or patients. For instance, Marchant et al. (2008) conducted a study on intermittent preventive treatment of malaria in pregnancy (IPTp) in Tanzania, where a researcher interviewed female respondents who had given birth to a child, as well as hospital employees, and then analyzed the coverage of IPTp by using the facilities at an antenatal clinic. During this survey, each piece of equipment and supplier were checked at the facility location to avoid delay during research.

According to Magnani et al. (2005) facilities such as drug treatment centers, HIV specialist hospitals, drug rehabilitation centers and gay bars are some good locations to find rare population involved in drugs and HIV activities. For example, for conducting a facility-based survey of HIV- infected people, first select a hospital with HIV specialist doctors where there is a higher possibility of getting a database of HIV infected people. This means that in the previous case hospitals (facilities) are being used to find and recruit target people. Subsequently, the researcher can conduct interviews at the facility location or can meet target samples personally at their home.

In spite of all of its advantages, the facility-based survey method has some bias problems such as volatile laws of the juridical system, stigmatization by society and police investigations (Turner et al. 2001). In addition, proper pre- survey management and skilled interviewers are needed for the study. Chopra et al. (2005) conducted a survey on intervention in a quality of care in Cape Town in which nurses were trained by local health services with the help of WHO‟s teaching modules. On top of that, prior juridical

permission was always required from the top authority of the hospital or government to perform the entire interview procedure at the facility location, which increases the time and cost of the survey. The number of special clinics for HIV- infected patients, especially in poor and developing countries, is limited, which places restrictions on the research (Shaghaghi, Bhopal and Sheikh 2011).

Furthermore, Turner et al. (2001) pointed out that the major issue for sampling and sample size is the rare occurrence of subjects at the facility place. For instance, target people who are unwilling (because of stigmatization) or unable (because of sickness or family problems) to come to the interview place creates a bias problem in the sampling. To overcome this rare occurrence problem, he suggested a strategy to conduct the entire survey at the facility location for five consecutive days, so that the maximum number of people would show up and participate in the research study.

2.4 Citizens Registry-Office Sampling

In empirical research, citizen‟s registry-office sampling is a typical method for surveying immigrants or minority populations. Relevant information fetched from the citizens registry-offices, which includes personal and contact information of the target population, can be used for sampling purposes. The main advantage of this sampling is that it is possible to collect the information of each target person who has registered themselves at the government office. Once the contact information list is available, interviews of the target population would be possible using RDD or direct face-to- face interview. However, with RDD there is a lower probability of the target sample being selected for an interview. Salentin (2007) describes three limitations that may occur in this sampling method:

a) In spite of mandatory regulations, some people do not register at the government office.
b) Sometimes people from minority groups do not update their address after changing their accommodation.
c) If the target population is living in rural areas, it is hard to achieve the expected sample size as the population is limited.

For accurate measurement of minorities in the specific location the researcher can extend our gravity model analysis using city registration methodology. The researcher can get contact information of the target population from schools with minority children. Even though this method is costly and time-consuming, there are better chances of finding people from minority groups near border locations.

2.5 Gravity Sampling

The basic idea behind the gravity sampling method is that the empirically based expectations about the prevalence of the target population at each location in their disseminated territory can be calculated on

the basis of the different attractiveness factors and distance to the specific location (Hoops and Schnapp 2013). This method can be used, in particular, when the target population is living in a relatively well-defined area of distribution. Initially the target area of a survey is so limited that such a high prevalence of the target population is achieved in the defined area and then telephone or face-to-face interviews would be beneficial to interview a person from the minority population. The technique is based on the multiplicative competitive interaction model (Nakanishi and Cooper 1974), which was designed as an extension of the univariate model by Huff (1964).

Gravity sampling requires confirmation of two elements. First of all we have to show that the probability of the target population increases at the residences which are located closer to the defined geographical area or respective country‟s border. Secondly, residences with a high attractive- ness for members of the target population should be selected more often than residences with less attractiveness.

If both hypotheses are correct, it can be considered that the distance to the border and the attractiveness of a place influence the choice of residence of target people. It makes it possible to accurately estimate the relative prevalence of the various destinations. Eventually, it can specify the number of interviews as well as the number of target population in the sample space for each location. In contrast to established methods like RDD, gravity analyses enable substantial cost savings when conducting a low prevalence survey (Hoops, Schnapp and Schaefer-Rolffs 2013).

The gravity probability depends on an empirical parameter called λ. The greater the impact of the distance on the prevalence, the greater is the value. In this paper, gravity analysis is tested with German minorities in Denmark, where both countries have a strong economy and share the same ethnicity. Nevertheless, λ depends on the sampling area and varies from country to country. The parameter that we used for the Danish-German border should be validated for other border region.

3. Methodological Analyses

3.1 Simulation Procedure

Based on an empirical study in Denmark, where information on the frequency of the German minority population was collected and 375 people who were living in Denmark and less than 100 kilometers from the Danish-German border were interviewed, we know that approximately 7.1% of all people feel that they belong to the German minority and around 8.2% of all households have at least one national minority. In the interests of simplification, we concentrated on the 14 most popular locations, which have a total population of 111,300 inhabitants.

Based on these data, we have randomly generated samples with this population size and assigned each case a household number, a household size, an age and the membership to the German minority (or not). Excluding this minority status, we have focused on the official statistics from Denmark (Danmarks Statistik 2013).

With reference to the work of Lu et al. (2009), we estimated the relevant network size of each national minority (see Figure 1).

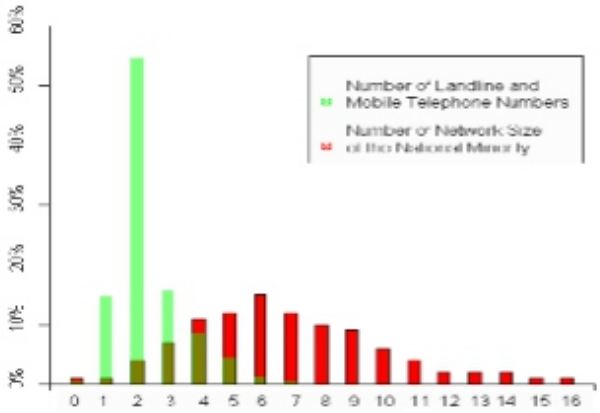


Figure 1: Assumed distributions of numbers and size

We assumed that our relevant size is similarly normally distributed and is slightly right-skewed with the maximum value of 16. About 99 percent of all national minorities should know at least one member who is not a part of theirhousehold. In that context more than half of the total minority is at least connected with six other German minorities.

Based on our Ipsos Democracy Pulse (see Hoops, Glantz and Michael 2013), we have used a RDD-similar technique for creating random mobile and landline numbers in Germany. Since there was no reliable data available before we expected that there would be no differences between these two countries. The assumed distribution of the sum of landline and mobile phone numbers of each person can be seen in Figure 2.

We have used $\lambda = 0.399$ for gravity sampling (Hoops, Schnapp & Schaefer-Rolffs 2013). Moreover, along with distance to the Danish-German border and also considered the election results of the minority party, the density of minority oriented organizations, schools and companies as relevant attractiveness variables. We supposed that our survey will constitute a public interest so we could use citizens registers to select a sample. This registers would be complete, correct and contain no duplicates. For facility-based sampling we restricted ourselves to German schools in Denmark so only people with children that go to school can be included.

We used bootstrapping to approximate the theoretical values of the bias, the cardinality of the sample and the variance of the inclusion probabilities. These statistics are computed by selecting sub-samples of the size of 100 people from the complete sample by different sampling techniques and repeated it 10,000 times. Therefore, the inclusion probability of a person in the sample can be estimated by the number of selections divided through the number of repetitions. While the bias is calculated by the relative medium deviation between sample and target population in some variables (age and number children), the cardinality of the sample space is measured as the number of different cases that are selected at least once divided through the number of target populations.

3.2 Results

In Table 2 you can see that the cardinality of the sample space is very low for facility-based sampling whereas gravity sampling has a complete sample space under our assumption of no missing minority in the register.

Table 2: Measures to compare different methods

Method	*Bias*	*Cardinality of the sample space*	*Variance of the inclusion probability*
FBS	63.10%	38.60%	25553.6
GS	9.50%	100%	127.5
RDD	9.40%	96%	10719.8
SS	17.40%	98.40%	3540.9

In RDD the cardinality of sample space is lower than in the gravity simulation as only people who have a landline or mobile connection can be interviewed. Because 1% of the total target population can‟t be recommended by any minority, the cardinality of the sample space in the snowball sampling simulation should be 99%. However, the cardinality is about 98.4%, so there is nearly one percent that could be selected, but is not. The facility-based method has the lowest cardinality among all sampling techniques, since only 3,153 out of 7,913 target people have school-going children.

On the contrary, facility-based sampling has the highest variance of the inclusion probabilities as compared to the others (see Figure 2), because there are many people who are eligible but have no chance to be part of the sample. The top- left plot shows only a few data points. Most of the people are selected between 310 and 330 times.

As we can see, gravity sampling produces the lowest variance of the inclusion probability and the selection frequencies are concentrated from 100 to 160 times. In contrast to RDD it is no household sample and enables roughly constant inclusion probabilities in the sample. The simulation of RDD shows a slightly larger variance with selection frequencies, which are lying in the range of 130 to 200 times. The more landline and mobile telephone numbers a minority member has, the greater the

probability of that person being selected in the sample space. So one minority has been randomly selected 862 times due to the possession of many telephone numbers in total.

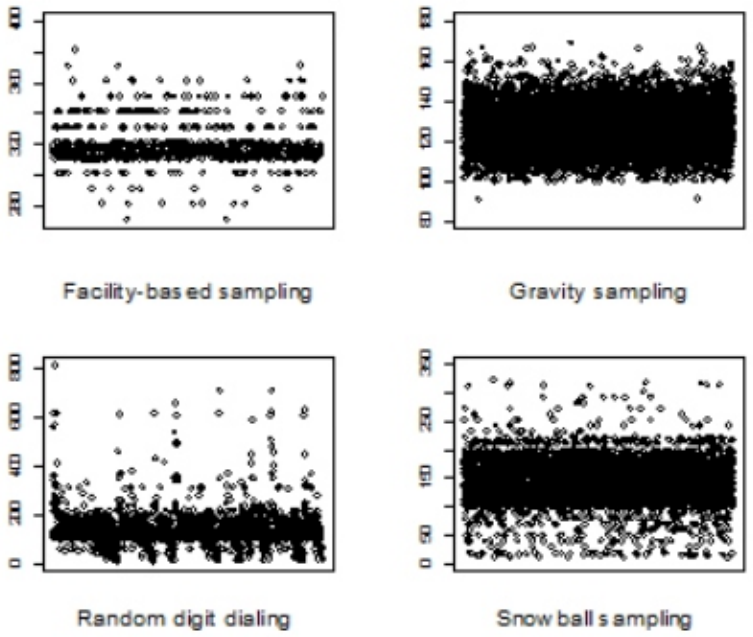


Figure 2: Plot of the inclusion frequencies of each technique

In snowball sampling the selection frequencies are concentrated between 100 to 200 times and a lower total variance of the inclusion probabilities is created. This is surely something surprising and calls our assumptions into question, because we expected a greater variance. The gravity sampling shows values from 91 to 169. Thus there is only less variation due to the expected frequency of about the average number of 126.37 (=(10,000*100)/7913) selections.

However, it is hardly surprising that all sampling methods create biased samples because of the very low sample size of 100. RDD and gravity samples show the lowest differences to the total population. So both methods are recommendable for representative studies. Furthermore, both biases should be asymptotically tending towards zero. Therefore, with snowball and facility-based sampling there can only be samples produced which differ greatly from the mean of the population. Because we even compute the distortion across the central portion of children in the household, the samples in facility-based sampling are greatly biased since the researcher can only reach target people with children. Snowball sampling is biased because older people tend to have a greater network size and are more often selected.

4. DISCUSSION AND CONCLUSION

Because of low prevalence, traditional survey methods such as random digit dialing and citizens registry-office sampling are very expensive when carrying out surveys of hard-to- reach populations. Therefore, alternative methods such as gravity or snowball sampling with some limitations would be favorable and recommendable.

Our simulations have shown some advantages of gravity sampling. First, this method in combination with citizens registers has the largest cardinality of the sample space under the assumption of correct official data. Second, there is no significant distortion detected and inclusion probability variation is far less. Although we received positive results for gravity sampling, the computations in real studies might behave significantly differently. Therefore, the gravity sampling method should be strongly tested in practice.

It is, however, possible that we are overestimating the variance of the inclusion probabilities in the RDD simulation, because the inclusion probability does not increase proportionally to the amount of telephone numbers. Surely a person who has ten mobile numbers will not always possess ten mobile phones and will not always be equally reachable at each number. This should be taken into account for later simulations.

It is also possible to incorrectly estimate the network size of national minorities. If, for example, a single individual knows more than 16 other members and/or the proportion of the minority without a network is considerably larger, this could lead to an increase in the inclusion probability in random samples. Nonetheless, our conservative assumption should not influence the conclusions.

We must of course admit that the calculation of bias using age and number of children has been established somewhat subjectively. Especially since both variables are obviously correlated with each other. But there was no valid multivariate marginal distribution that we could have used for our simulations. It is also impossible to judge the value of the estimation of loading parameters lambda. So it only remains to be recommended that an appropriately practical method test should be conducted with a large sampling to calculation the deviations from official statistics.

REFERENCES

[1] Agadjanian, V. and Zotova, N. (2012) ,,Sampling and Surveying Hard-to-Reach Populations for Demographic Research: A Study of Female Labor Migrants in Moscow", Demographic Research, 26(5) 131-150.

[2] Atkinson, R. and Flint, J. (2001) ,,Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies", Social Research, 33(1) 1-4.

[3] Berg, S. (1983) ,,Random Contact Processes, Snowball Sampling and Factorial Series Distributions", Journal of Applied Probability, 20(1) 31-46.

[4] Berg, S. (1988) ,,Snowball Sampling", Encyclopedia of Statistical Sciences, 8, 529–532.

[5] Biernacki, P. and Waldorf, D. (1981) ,,Snowball Sampling: Problems and Techniques of Chain Referral Sampling", Sociological Methods and Research, 10(2) 141-163.

[6] Brunner, J. A. and Brunner, A. G. (1971) ,,Are voluntarily Unlisted Telephone Subscribers Really Different?", Journal of Marketing Research, 8(1) 121- 124.

[7] Chopra, M., Patel, S., Cloete, K., Sanders, D. and Peterson, S. (2005) ,,Effect of an IMCI Intervention on Quality of Care across four Districts in Cape Town, South Africa", Archives of Disease in Childhood, 90(4) 397-401.

[8] Danmarks Statistik (2013) ,,StatBank Denmark" [electronic resource]. Copenhagen: Statistics Denmark. http://www.statbank.dk.

[9] Esslemont, D., Petersen, S. and Selvakumar, K. S. (1992) ,,Telephone Directories as Sampling Frames", Marketing Bulletin, 3(5) 38-45.

[10] Ezzati-Rice, T. M., Frankel , M. R., Hoaglin, D. C., Loft, J. D., Coronado, V. G. and Wright, R. A. (2000) ,,An Alternative Measure of Response Rate in Random-Digit- Dialing Surveys that Screen for Eligible Subpopulations", Journal of Economics and Social Measurement, 26, 99- 109.

[11] Faugier, J. and Sargeant, M. (1997) ,,Sampling Hard-to- Reach Population", Journal of Advanced Nursing, 26(4) 790-797.

[12] Frankel, M. R., Battaglia, M. P., Link , M. and Mokdad, A. H. (2007) ,,Integrating Cell Phone Numbers into Random Digit-dialed (RDD) Landline Surveys", Proceedings of the Survey Research Methods, 3793- 3800.

[13] Fuchs, M. (1994) Umfrageforschung mit Telefon und Computer. Weinheim: Beltz Psychologie Verlags Union.

[14] Glasser, G. J. and Metzger, G. D. (1972) ,,Random Digit Dialing as a Method of Telephone Sampling", Journal of Marketing Research, 9, 59-64.

[15] Groves, R. M. and Lyberg, L. E. (1988) ,,An Overview of Non Response Issues in Telephone Surveys", in: Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholis, W. L. and Waksberg, J. (eds.). Telephone Survey Methodology. New York: Wiley.

[16] Heckathorn, D. (1997) ,,Respondent-driven sampling: A New Approach to the Study of Hidden Populations", Social Problems, 44, 174-199.

[17] Hendricks, V. and Blanken, P. (1992) ,,Snowball Sampling: Theoretical and Practical Considerations, in: Hendricks, V. and Blanken, P. (eds.). Snowball Sampling: A Pilot Study on Cocaine Use. Rotterdam: IVO.

[18] Hoops, C., Glantz, A. and Michael, T. (2013) ,,Innovations in Rolling-Cross-Section-Design", Journal of Political Science, 23(2) 215-243.

[19] Hoops, C. and Schnapp, K.-U. (2013) ,,Sampling of Hard-to-Reach-Populations", Paper presented at European Survey Research Association Conference, Ljubljana, July 15-19.

[20] Hoops, C., Schnapp, K.-U. and Schaefer-Rolffs, A. (2013) Zur Ziehung von Stichproben in schwer erreichbaren Zielgruppen mit Gravitations-analytischen Methoden. Hamburg: University of Hamburg (WiSo-HH Working Paper Series; WP-2013-4).

[21] Huff, D. L. (1964) ,,Defining and Estimating a Trading Area", Journal of Marketing, 28(3) 34-38.

[22] Illenberger, J., Flötteröd, G. and Nagel, K. (2008) An Approach to Correct Biases Induced by Snowball Sampling. Berlin: Technical University Berlin (VSP Working Paper 8-16).

[23] Jäckel, E. (1959) Die Schleswig-Frage seit 1945: Dokumente zur Rechtsstellung der Minderheiten beiderseits der deutsch-dänischen Grenze. Frankfurt am Main: Metzner.

[24] Johnston, L. and Sabin, K. (2010) ,,Sampling Hard to Reach Populations with Respondent Driven Sampling", Methodological Innovations Online, 5(2) 38-48.

[25] Kendall, C., Kerr, L. R., Gondim, R. C., Werneck, G. L., Macena, R. H., Pontes, M. K., Johnston, L. G., Sabin, K. and McFarland, W. (2008) ,,An Empirical Comparison of Respondent-Driven Sampling, Time-location Sampling, and Snowball Sampling for Beha-vioral Surveillance in Men Who Have Sex with Men, Fortaleza, Brazil", AIDS and Behavior, 12(1) 97-104.

[26] Lopes, C. S., Rodrigues, L. C. and Sichieri, R. (1996) ,,The Lack of Selecton Bias in a Snowball Sampled Case- Control Sudy on Drug Abuse", International Journal of Epidemiology, 25(6) 1267-1270.

[27] Lu, Y.-E., Roberts, S., Lio, P., Dunbar, R. and Crowcroft, J. (2009) „Size Matters: Variation in Personal Network Size, Personality and Effect on Information Transmission", Computational Science and Engineering, 4, 188-193.
[28] Magnani, R., Sabin, K., Saidel, T. and Heckathorn, D. (2005) „Review of Sampling Hard-to-Reach and Hidden Populations for HIV Surveillance", AIDS, 19(2) 67-72.
[29] Marchant, T., Nathan, R., Jones, C., Mponda, H., Bruce, J., Sedekia, Y., Schellenberg, J., Mshinda, H. and Hanson, K. (2008) „Individual, Facility and Policy Level Influences on National Coverage Estimates for Intermittent Preventive Treatment of Malaria in Pregnancy in Tanzania", Malaria Journal, 7, 260.
[30] Marpsat, M. and Razafindratsima, N. (2010) „Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue", Methodological Innovations Online, 5(2) 3-16.
[31] Massey, J. T. and O'Connor, D. (1997) „Response Rate in Random Digit Dialing (RDD) Telephone Survey" Paper presented at American Statistical Association Conference, Anaheim, August 10-14.
[32] McKenzie, D. J. and Mistiaen, J. (2009) „Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys", Journal of the Royal Statistical Society 172(2) 339-360.
[33] Miller, K. W., & Bilder, L. B., Stillman, F. A. and Becker, D. M. (1997) „The Feasibility of a Street- Intercept Survey Method in an African-American Community", American Journal of Public Health 87(4) 655-658.
[34] Nakanishi, M. and Cooper, L. G. (1974) „Parameter Estimation for a Multiplicative Competitive Interaction Model - Least Squares Approach", Journal of Marketing Research 11(3) 303-311.
[35] Salentin, K. (2007) „Die Aussiedler-Stichproben- ziehung", Methoden - Daten - Analysen 1(1) 25-44.
[36] Salganik, M. and Heckathorn, D. (2004) „Sampling and Estimation in Hidden Populations Using Respondent- Driven Sampling", Sociological Methodo-logy, 34, 193- 239.
[37] Sangngam, P. and Suwattee, P. (2010) „Stratified Inverse Sampling", Journal of Thai Statistical Association, 10(1) 69-86.
[38] Shaghaghi, A., Bhopal, R. S. and Sheikh, A. (2011) „Approaches to Recruiting "Hard-To-Reach" Populations into Research: A Review of the Literature", Health Promotion Perspectives, 1(2) 86-94.
[39] Turner, A. G., Angeles, G., Tsui, A. O., Wilkinson, M. and Magnani, R. (2001) Sampling Manual for Facility Surveys: For Population, Maternal Health, Child Health and STD Programs in Developing Countries. Chapel Hill: University of North Carolina (Measure Evaluation Manuel Series; Working Paper 3).

Author Profiles

Christian Hoops is a German statistician who works for Ipsos Public Affairs. In 2011 he received the EHI science prize for innovative methodology development, after which he moved to the Technical University of Dortmund as a lecture of quantitative methods and scientific work. His main research priority focuses on sampling, survey design and electoral research.

Rahul Pathare holds different certificates of industrial electronics, engineering and technology. He has 3 years of work experience in academics as an assistant professor at the Narsee Monjee Institute of Management Studies at the University of Mumbai. His area of focus has been business, statistics, scenario planning & technology.

A Survey on Cluster Head Selection Techniques

Monika[1], Pooja Ahlawat[2]
[1]M.Tech Student, R. N. College of Engineering & Management,
Maharshi Dayanand University, Rohtak, Haryana, India
[2]Associate Professor, Department of Computer Science and Engineering, R. N. College of Engineering & Management, Maharshi Dayanand University, Rohtak, Haryana, India

ABSTRACT

Low energy adaptive Hierarchical clustering Protocol (LEACH) is used to transfer the data to sink node by utilizing minimum amount of energy. It forms clusters of nodes that can adapt according to the various parameters like energy etc. TO enhance the network lifetime various modifications are proposed to the LEACH. These modifications include different procedures to form clusters and to select cluster head. This paper compares LEACH and its various enhancements. This paper also describes the active and the passive clustering.

Keywords: Wireless Sensor Network, Clustering, LEACH, E-LEACH, C-LEACH, LEACH-A, LEACH-F

1. INTRODUCTION

Wireless sensor networks are usually consists of thousands of inexpensive, low-powered sensing devices having limited battery and communication resources [1,2]. These networks offer a good range of applications in the military as well as in civilian applications. The most signification feature of the WSN is the low deployment cost that results in several limitations like limited battery life etc. The network life can be prolonged by saving the battery. Therefore, in order to reduce the power consumption of wireless sensor networks, several mechanisms are proposed such as control packet elimination, topology control, and data aggregation [3]. Data aggregation targets to combine and summarize data packets of several sensor nodes so that amount of data transmission is reduced [1].

2. CLUSTERING

All nodes can transmit their data to the sink node but it will increase the network traffic So to decrease the network traffic i.e. to reduce the energy consumption the several Clustering technique are used. In these techniques a group of nodes is formed and they select a Cluster Head (CH) for transmission. All nodes within the cluster transmit their data to CH, where, it aggregates data and send to the Base Station (BS). Now the cluster head will transmit at large distance so, less energy is consumed. A clustered protocol shows better performance in terms of energy consumption when compared to other protocols. Clustering networks are classified in two types i.e. homogeneous and heterogeneous network [4]. All

clustering techniques consist of two phases; setup phase and steady state phase. In setup phase, formation of clusters and election of CHs is performed and in steady state phase, nodes transmit data to CH and it aggregates the data for sending to BS.

Clustering techniques are of three type active, passive and hybrid. In active clustering scheme, all the sensor nodes are synchronized to maintain the clusters. While in the passive technique, no control packets are used. It exploits the data packets to transmit neighbor's information. Hybrid approaches use a combination of active and passive techniques [5].

2.1 Active Clustering

In this technique hello packets are used to collect information about the network. Active clustering algorithms use various criteria for the selection of a cluster- head Lowest-ID and use the identifiers of nodes and the number of neighbors. Basu et al.[6] adds the degree of mobility to the LEACH. It also assigns different roles to distinct nodes according to the Round-Robin policy management. Bagrodia et al.[7] said these algorithms require two phases: Neighbor discovery and cluster formation phase. However, nodes are assumed fixed over the steps and synchronization between them is necessary for the success of these algorithms. In addition, following each change of network topology these steps are repeated periodically, which degrades the stability of clusters.

2.2 Passive Clustering PC

Passive clustering [8] is the demand cluster formation protocol that does not use any protocol-specific control packets. In this data packets are used to transmit neighbor's information. Clusters are formed flooding the data message. It reduces the initial set-up period that results in reduction of the total energy consumed as the main function of the clusters is to optimize the exchange of flooded messages. Passive Clustering uses the MAC frame to encode the state of a network node. Passive Clustering uses two bits to encode four states (1) Initial, (2) Cluster head, (3) Gateway and (4)
Ordinary [5].

3. LEACH AND ITS DESCENDANT

3.1 LEACH

Low Energy Adaptive Clustering Hierarchical Protocol (LEACH) uses the following techniques to achieve the design goals: randomized, self-configuring and adaptive cluster formation, Local control for data transfers and low-energy media access control and application specific data processing [11]. LEACH protocol has many rounds and each round has two phases, a setup phase and steady state phase,

in set up phase it provides cluster formation in adaptive manner and in the steady state phase transfer of data takes place. LEACH uses a TDMA or a CDMA MAC to reduce inter-cluster and intra cluster collisions. Cluster formation based on many properties such as the number and type of sensors, communication range and geographical location. The energy consumption of the information gathered by the sensors node to reach the sink will depend on the number of cluster heads and radio range of different algorithms, because the energy consumption can be reduced by organizing the sensor nodes in the clusters [12].

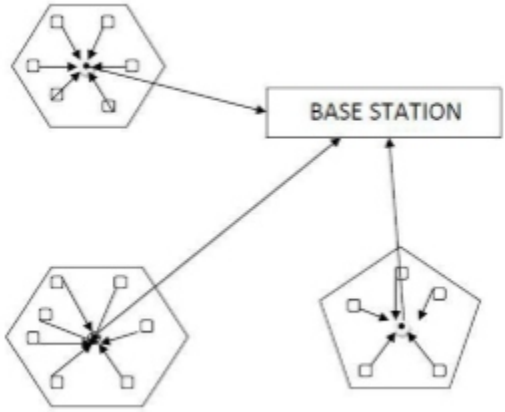


Figure 1: Leach protocol architecture [13]

Although LEACH protocol prolongs the network lifetime in contrast to plane multi-hop routing and static routing, it still has problems. The cluster heads are elected randomly, so the optimal number and distribution of cluster heads cannot be ensured. The nodes with low remnant energy have the same priority to be a cluster head as the node with high remnant energy. Therefore, those nodes with less remaining energy may be chosen as the cluster heads which will result that these nodes may die first. The cluster heads communicate with the base station in single-hop mode which makes LEACH cannot be used in large-scale wireless sensor networks for the limit effective communication range of the sensor nodes [13].

3.2 Enhanced-leach (E-LEACH)

E-LEACH basically removes overload energy consumption problem of the LEACH. The E-LEACH adopts the same round concept with the original LEACH. E-LEACH selects optimal number of cluster-heads to enhance the performance. If the number of cluster-heads is less then each cluster-head covers larger region, this will lead the problem that some cluster-members get far from their cluster-heads and consume much more energy. When the numbers of cluster heads are less then more energy is consumed in the communication between cluster heads and the base station. Therefore, it is necessary to select optimal cluster head number to make the energy consumption minimum. In the E-LEACH minimum spanning tree between cluster heads is used. The node with largest residual energy is selected as the root node [13].

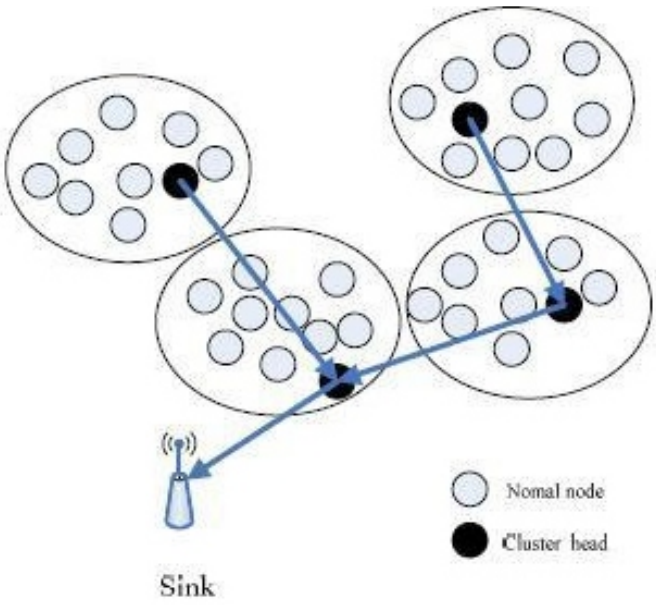


Figure 2: Architecture of E-LEACH [13]

3.3 LEACH-C (Centralized Low Energy Adaptive Clustering Hierarchy)

It involves a centralized clustering algorithm. The steady state of LEACH-C will remains same as LEACH whereas the setup phase is different. In the setup phase of LEACH-C each node within the network transfers the information about the current location and energy level to the base station. The base station uses this information to produce better clusters that requires the less energy for data transmission [14].Each sensor node is equipped with GPS to track the location. The base station elects the cluster head on the basis of energy level of node. Leach-C has a deterministic threshold algorithm to form the cluster and to elect the cluster head [14].

3.4 LEACH-F

LEACH-F, here F stands for the fixed. In this technique the number of clusters remains fixed throughout the network lifetime and the cluster heads rotated within its clusters. As the Steady state phase forms the clusters so the steady state phase of LEACH-F is similar to that of LEACH. LEACH-F can provide the energy savings but doesn't support flexibility. In other words LEACH-F doesn't provide mechanism to add or remove sensor nodes from the sensor networks [11].

3.5 LEACH-A (Advanced Low Energy Adaptive Clustering Hierarchy)

LEACH-A decreases the head node energy consumption to increase the network lifetime. In this algorithm the data is processed using mobile agent technique based on Leach. Advanced Leach uses a

synchronized clock to decrease the node's failure probability hence increasing the time interval before the death of the first node [10].

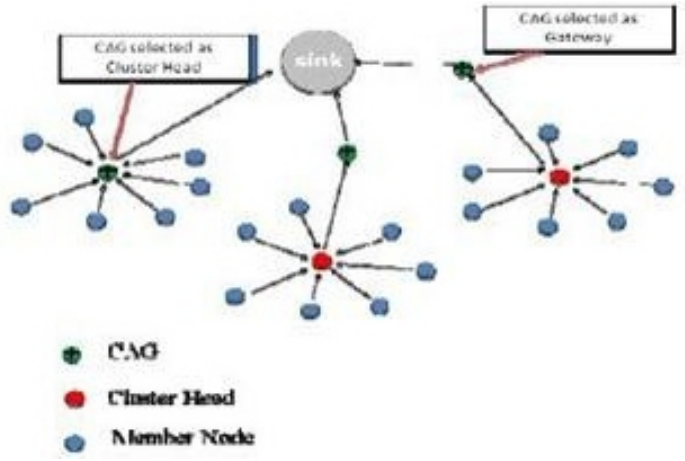


Figure 2: LEACH – A Network style

If n is the total number of nodes in network and m being the fraction of n that are equipped with clock. Then these nodes are called CAG nodes, the nodes selected as cluster heads and the others (1-m) x n as the normal nodes.

In this paper, a well-known protocol in WSN called LEACH is described. We have surveyed different clustering algorithms in wireless sensor networks along with LEACH. We have presented the comparison of different LEACH descendant. Each of the routing protocol has its own advantages compared to the fundamental leach routing protocol. The drawbacks and issues addressed by the LEACH protocol are also discussed. We have found that the some energy efficient algorithms increase the network lifetime. In future we propose to modify the LEACH by using soft computing techniques to enhance the network lifetime. As soft computing techniques are highly dynamic so by applying these techniques LEACH can adapt easily according to parameters changes results in energy saving.

Table 1: Comparison of LEACH and its DESCANDS

Clustering Routing Protocol	Classification	Mobility	Scalability	Self-organization	Randomized Rotation	Distributed	Centralized	Hop count	Energy efficiency	Homogenous	Data aggregation
LEACH	Hierarchical	Fixed BS	Limited	Yes	Yes	Yes	No	Single Hop	High	Yes	Yes
E-LEACH	Hierarchical	Fixed BS	Very Good	Yes	Yes	Yes	No	Single Hop	Very High	No	Yes
LEACH-C	Hierarchical	Fixed BS	Good	Yes	Yes	No	Yes	Single hop	Very High	Yes	Yes
LEACH-F	Hierarchical	Fixed BS	Limited	No	Yes	No	Yes	Single hop	Very High	Yes	Yes
LEACH-A	Hierarchical	Fixed BS	Good	Yes	Yes	Yes	No	Single hop	Very high	No	Yes

REFERENCES

[1] *Suat Ozdemir et al., "Secure data aggregation in wireless sensor networks: A comprehensive overview", 2009 Elsevier B.V. All rights reserved.*

[2] *J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Networks 52 (12) (2008) 2292–2330.*

[3] *K. Akkaya, M. Demirbas, R.S. Aygun, The Impact of Data Aggregation on the Performance of Wireless Sensor Networks, Wiley Wireless Commun. Mobile Comput. (WCMC) J. 8 (2008) 171–193.*

[4] *Javaid, Nadeem, et al. "On Sink Mobility Trajectory in Clustering Routing Protocols in WSNs." arXiv preprint arXiv: 1304.0984 (2013).*

[5] *Aouad Siham et al., "ADVANCED ENERGY EFFICIENT PASSIVE CLUSTERING MOBILITY IN WIRELESS SENSOR NETWORKS", American Journal of Applied Sciences 10 (12): 1558-1569, 2013 ISSN: 1546-9239 ©2013 Science Publication doi:10.3844/ajassp.2013.1558.1569 Published Online 10 (12) 2013 (http://www.thescipub.com/ajas.toc)*

[6] *Basu, P., N. Khan and T.D.C. Little, 2001. A mobility based metric for clustering in mobile ad hoc networks. Proceedings of the 21st International Conference on Distributed Computing Systems, Apr. 16-19, Meas, AZ., pp: 413-418. DOI: 10.1109/CDCS.2001.9187738*

[7] *Bagrodia, R., R. Meyer, M. Takai, Y. Chen and X. Zeng et al., 1998. Parsec: A parallel simulation environment for complex systems. Computer, 31: 77-85. DOI: 10.1109/2.722293.*

[8] *Gerla, M., T.J. Kwon and G. Pei, 2000. On-demand routing in large ad hoc wireless networks with passive clustering. Proceedings of the Wireless Communications and Network Conference, Sept.23-28, IEEE Xplore Press, Chicago, IL., pp: 100-105. DOI: 10.1109/WCNC.2000.904608.*

[9] *K.Ramesh et al., "A COMPARATIVE STUDY OF CLUSTERHEAD SELECTION ALGORITHMS IN WIRELESS SENSOR NETWORKS", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.4, November 2011.*

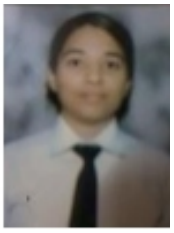
[10] *W.Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", In Proc. 33rd HICS, 4-7 Jan, 2000, Vol 2, pp. 10.*

[11] *Vinay kumar et al., " Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011 ISSN (Online): 1694-0814 www.IJCSI.org.*

[12] *H. Junping, J. Yuhui and D. Liang "A Time-based Cluster-Head Selection Algorithm for LEACH" IEEE Symposium on Computers and Communications, 2008, pp.1172-1178.*

[13] *Ravneet Kaur et al., "Comparative Analysis of Leach and Its Descendant Protocols In Wireless Sensor Network", International Journal of P2P Network Trends and Technology- Volume3Issue1- 2013.*

Author Profile



Monika received the B.Tech and M.Tech degrees in Computer Sciece and Engineering from R. N. College of Engineering & Management in 2013 and 2015, respectively.

Does the Database Functional Dependency and its Normalization Make Uniform Database Management in Future?

Yagyanath Rimal
Lecturer,
Computer Science/IT Pokhara University

ABSTRACT

This review deals with the fundamental concept of database management, the functional dependence of data, and its normalization, to organize daily records management for information processing. The primary purpose is to explain the functional dependence its normalization using various suitable real world example which only meet present demand of database organization. Therefore, this paper presents the different database styles to meet the data management system for future analysis.

Keywords: Database, Management Systems, Relational, Key, Primary, Candidate, N, FBCNF, ER Model.

INTRODUCTION

Today database management systems have become an integral part of all types of work, whether in the management of business records or daily data organization. Data has not only become new thought but it is also the catalyst that is driving organizations to new heights of success (McKendrick, 2017). Similarly, blog atlas system.com, argued that the database management is developing new trends that focus on security, mechanization, cloud and NoSQL were current trend of data design management. Similarly (Lohr, 2016)explained in New York survey claimed that 80% of the time and effort of a data scientist is dedicated to collecting, cleaning and preparing data for analysis as data sets in various sizes and of a different nature. Similarly recent survey of 16,000 data professionals concluded that the mostcommon challenges to data science included dirty data (36%), lack of data science talent (30%) and lack of management support (27%) are the essential component of database management. Most data professionals concludes that there were five categories of challenges grouped from 20 common changes(Hayes, 2018). Although, data mining is considered as an interdisciplinary sub domain of computer science, which basically apply intelligent methods in order to extract data in a fixed patterns that ultimately used for analysis for future planning is known as data mining. The data collected from different applications Knowledge Discovery in Databases (KDD), often called data mining, which aims to discover meaningful information from large collections of raw data (Sangeeta, 2016). Data analytics will be the closely integrated platforms are essential to identify patterns, optimize behavior and detect

anomalies in internet of thing (IoT) without human intervention (Markarian, 2017).However the recent trendsof database management focus on join SQL / NoSQL. Therefore databases in the cloud / Platform as a service, automated management,which sets greater focus on security that always requires data stored in predefined patterns so that we could easily meets in the future hazard of data organization (Gates, 2017). (FRY, 1976) The data science has explained sharply increased in size in recent years since 1950 when McGee discusses the successful of generation. Human needs and requirements are increasing in the modern electronic age with the high use of electronic devices and information is stored for future requirements are the side of database management. If we analyze the electronic information storage curve, the use of information by the administration for future analyzes has been much broader per year; which finally makes human life easier than in recent years. The records of storage, organization and retrieval of information and its extraction procedures are the first secondary management for the data scientist. (Peng, 2017)He estimated that the number of mobile phone users in the world is expected to exceed the five billion in 2019. In 2016, it is estimated that 62.9 percent of the world's population already had a mobile phone. The penetration of mobile telephony is expected to continue growing, rounded up to 67 percent in 2019. However, most of the growth of the mobile phone market can be attributed to the growing popularity of smartphones. Around 38 percent of mobile devices were smartphone users.

Similarly database is an organized collection of data typically stored in electronic format that allows you to quickly insert, manage, organize and retrieve data .A present, data scientist are planning better algorithms to get and the future problem of the database. The data cleanup plays important role in data management in the right direction (Shakir Khan, 2012). Although, data scientists are investigating ways for storing data through quantum physics, too little data connected to the spin of an electron (Pinola, 2017). The rapid advances in storage, communications and processing allow us to move all information to cyberspace (Gray, 2017).Mark Whitehorn, emeritus professor of analysis at university of Dundee, supports this multidimensional, business intelligence (MBI), data storage, hadoops helps bring big data at data warehousing, dimensional and factual tables, data warehouse performance management through R programming (Whitehorn, 2017).Gray also faced the most challenging problem is understanding data that there is no doubt that most of the data will be online soon, since it is economical to store data on computers and data server computers in cloud. Organizing these huge data files so that people can easily convert into information is the real challenge which should be based on data patterns, trends, anomalies and relevant information from a large database is most interesting areas of data management (U.M. Fayyad, 1995).

Similarly, the survey conducted by Gil Press revealed that 57% of the data scientific believe that the cleaning and organization of data is the most tedious and least pleasant task in the data science process and 19% believe that the collection of sets of data is the least pleasant task (Press, 2018). Data becomes information once processed, which includes the identification of data analysis problems that offer the best opportunities for the correct organization of data sets and variables, the collection of large-scale structured and unstructured data from different sources, validate data to ensure accuracy, integrity and consistency, apply models and algorithms to exploit large data stores, analyze data to identify patterns and trends, interpret data to discover solutions and opportunities, communicate results to stakeholders through graph visualization. Therefore, there is a great deal of data that can be analyzed to understand the past or predict the future and support decision making. Although analyzing data by hand takes a long time. It is found that data has become much more prevalent, as business leaders now realizing that a data backed approach is the only way to the clients or investors(Ciupa, 2018).Data scientist's predictive analysis can be as good as the data collected for the stored values. The management of database systems in a systematic way is very necessary for database design to meet future demand.

The data in the table with rows and columns, is known as records designed in an organized way. The organization of the records in the column must be based on other records that must exactly match the organization's policies and the data structures. Suppose a relational table t1 [a] = t2 [a] and t1 [b] = t2 [b] are always uniform, here t1 and t2 are rows of tables and a and b are its column attributes. Therefore, functional dependency tries to keep the primary keys so that each row of data has a unique relationship with other data from another table, while both dependencies must be of the same mode. Therefore, data management requires a secondary priority for normalization and hence functional dependency. Likewise, data in the spreadsheet is stored in rows and columns which indirectly known rows, and columns are known as single-argument attributes. Where each row in the table corresponds to individual records and contains different attributes that describe the row. If the records of a single table become insufficient, then other records must be stored in the database, so a relational database is required, which can be of different types and models to handle all records. Therefore, a relational database (RDBMS) is a collection of data tables that describe and organize mainly a relational model. Each table must identify a column or a group of columns called the primary key column. Then, only the results of the predictive analysis were assembled to the stored values. The database administrative system maintain of interrelated data and a set of programs to be stored and accessed. The management of technical data systems manager maintain necessary for the operation of data where information history is available in an organized form for functional dependency should be the only data column that must be unique so that each line has the similar data value. Therefore, data management requires a priority to evaluate functional dependency of all attributes, which contribute relational database collection of tables called the primary key column.

Functional Dependency

Table 1: Student Database					
Rollno	**Name**	**Phone**	**State**	**Country**	**Age**
1	Rupa	98467329	one	Nepal	21
2	Rupa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22

Functional dependency is a restriction between two sets of attributes in a database relationship which is indicated by an arrow (→) among various data attributes. If an attribute A functionally determines B, then it is written as a → B.For example, employeeid → name means that employeeid functionally determines the employee's name. As another example in a student table database, {studentid, time} → {lectureroom}, the student's identification and time determine the lecture room where the student should be appears on time. The Rollno field has unique values, known as primary key fields, each student has a unique roll number in any classes. The fields have Rollno-> Name, Rollno-> Phone, Rollno-> State, Rollno-> Country, Rollno->Age are various relational properties. The State and Country field still have similar values, so they can have another relationship.

State->Country. The fields in the country also have a similar data value known as a set of functional dependencies. The nearest joint attributes can be determined by the nearest attributes. Therefore, it could be a table format of two relationships like (Rollno) + = {Rollno, Name, Phone, State, Country, Age} with another type relationship whereas (State +) = {State, Country} has another type. The attribute sets of all attributes have a super key here Rollno already assumes a unique value and each student table also has a unique name where, as if no subset of this relationship can functionally determine, all attributes will be candidate key here (Rollno +) is a candidate key whereasRollno and Name are super keys. Thus, the attributes that are part of the candidate key are known as main attributes while the others are non-core attributes. The canonical coverage occurs when any database is updated should verify the functional dependency when changes to the new database system should be restored with the help of the same functional dependency of the original set. Theextraneous attributes occur when functional dependence is extraneous if we can eliminate it without modifying the closure of the set of functional dependencies. Which is mainly based on the assumption that every left side of a functional dependency is unique.

Normalization

Database normalization is the process of organizing database attributes to reduce or eliminate data redundancy (with the same data but in different locations) that unnecessarily increases the size of the database as it grow. The same data are repeated in many places. Incoherency issues arise during insert,

delete, and update operations because data is spread across multiple files and files can be in different formats; it is difficult to write new application programs to retrieve the appropriate data. Data redundancy eliminated from Table no 2, so data consistency and integrity are increased so that data is secure and flexible for database recovery. Standardization is the process of organizing data in the database that includes creating and establishing relationship between tables used to eliminate redundancy with repeating a group, not independent of the key, the independent multiple relationship. There are some forms of normalization that eliminate redundancy.

Table no 2:Un-Normalized Database				
Roll	**Name**	**Courses1**	**Course2**	**Course3**
105	Sila	Java	Oracle	Oracle
106	Binod	C++	Java	Computer

First Normal Form

The first normal form indicates that the data is in entity form, which means that the following conditions have been met. If the relationship contains compound and multivalued attributes, it violates the first normal form. If all attributes are in the unique value attribute, which eliminates group repetition, the individual tables that eventually create a separate table for each related data set are configured with a unique primary key for the data set. While the design of the main principle of decomposing the main table into minor is that it sets multiple fields in a single table to store similar data, and each cell contains an atomic number, each column must have a single table header.

Table no 3: 1st Normalized		
Roll	**Name**	**Course**
105	Sila	Java
105	Sila	Oracle
105	Sila	Oracle
106	Binod	C++
106	Binod	Java
106	Binod	Computer

In the same way table no 4 could be normalized with attributes of Roll no has two phones that can handle the repetition of all the attributes in a separate tuple in the table data. Similarly the table no 2 can be restructure as table no 3 with first normal form by applying the repetition of roll no and name course properties.

Table no 4: Student database					
Rollno	**Name**	**Phone**	**State**	**Country**	**Age**
1	Rupa	98467329	one	Nepal	21
		98456432			
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22
1st Normalization Form					
Rollno	**Name**	**Phone**	**State**	**Country**	**Age**
1	Rupa	98467329	one	Nepal	21
1	Rupa	98456432	one	Nepal	21
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31

Third Normal Form

The 3NF checks for transitive dependency (second level dependency of primary key) which eliminates field that's do not depends on the key values that are not records key do not belong in the table. If the component of a group of field apply to move single records put these field in separate table. Suppose R(a,b,c,d) is relational table having ab->c and c->d relation then R(a,b,c) and R(c,d) which are non-triviali.e. a->d has transitive dependency of second level with c column.

Roll	Name
105	Sila
105	Sila
105	Sila
106	Binod
106	Binod
106	Binod

Student table

Table no 8 : Normalized Student

Roll	Course	Credit	Marks
105	Java		
105	Oracle		
105	Oracle		
106	C++		
106	Java		
106	Computer		

Registration table

Table no :9 Another database

Name	Faculty	Department
Sila		
Bindo		

Faculty table

depends on the key values that are not records key do not belong in the table. If the component of a group of field apply to move single records put these field in separate table. Suppose R(a,b,c,d) is relational table having ab->c and c->d relation then R(a,b,c) and R(c,d) which are non-triviali.e. a->d has transitive dependency of second level with c column.

The first table no 3 can be decomposed as three above tables when organization needs more attributes to be required. Likewise the above table Roll no -> Name, Phone, State, Country, Age which violate Roll no -> State, State -> Country third Normal form so we may easily re-structure as Student (Roll no, Name, Phone, State, Age) and State (State, Country) tables in separate format for more attributes.

Table no 10: Normalized Student Database

Rollno	Name	Phone	State	Age
1	Rupa	98467329	one	21
2	Rupa	98456432	one	21
3	Rupaa	98467392	two	23
4	Sunita	98467329	one	31
5	Bimal	98467320	four	22

Rollno	State	Country
1	one	Nepal
2	one	Nepal
3	two	Nepal
4	one	Nepal
5	four	Nepal

Referential integrity is a database concept use to ensure the relationship between many tables in database remains synchronized during data modification. R1 can be used to ensure data is class may be

helpful in optimizing your database environment and can assist in early detection of errors. The combination of primary key and foreign key constraints can be used to enforce referential integrity of any database. In addition foreign key referencing to primary key can also reference. A uniqueness of constrain help to maintain referential integrity which sets some special reference so that it can easily locate the particular value inside table.

If you want to get data from table 11, it is very difficult to get the appropriate value of c column because this table has no row references. Therefore, special keys are needed to access special data from table no 11. There are several types of keys that help relationships to find a special value in table no11. The field b and c are not the key from which it cannot search for special values from the table. Therefore, a->bc and bc-> a are the key, but if c-> a is not a key whose value cannot predict each of the attributes in the table. The super keys are those keys that can contain one or more attributes through which it can easily identify the row and column separately. Suppose that in table R (a, b, c, d) has some functional dependency a-> bcd and ab ->cd, abc-> d, bd-> abc-> to all the others except c-> ab we can easily find all the attributes become super key that could not find the value of b. The candidate key is the minimum key, only the super key. If you apply the concept of the appropriate subset of the super key must have minimum values it is known as a candidate key. The bd- > ab attribute is also a candidate key because it has been left in relation. The primary key is a candidate key in a single pass that is easily selectable by the database administrator. It is known as a primary key that must be only one primary key in any table.

Table no 11:

a	b	c	d
1	a	x	
2	b	x	
3	b	x	
4	c	x	

Similarly, another relation R (A,B,C,D,E,F,G,H) table where AB-> C, A-> DE, B-> F, F-> GH clearly shows that AB does not have an inbound edge, therefore, AB are essential attributes, when designing tables A and B are always essential attributes through which we can easily find other attribute values in relational table.

The another example of R(A,B,C,D,E,F,G,H) relational table has AB->C, BD->EF,AD->G,A->H relation from relational inbound analysis ABD become candidate key through which it could find all values of attributes. In sometime it is necessary to find out the conditional relationship of attributes through which it could easily find out all attributes. Suppose R(A,B,C,D,E) having relationship AB->CD, D->A and BC->DA here B doesn't have incoming edge but could not be single candidate key therefore it is better to combine AB,BC,BD might be candidate key of tabular relationships.

Second Normal Form

The records in the relational table must be first in 1st normal form and must not contain any partial dependency. The non-prime attributes other than candidate key create separate tables for sets of values that apply to multiple record. The records of new table has with a foreign key and primary key. The records should not dependent on anything other than a table's primary key including compound key necessary. If a table of relation R(a,b,c,d) and ab->d and b- >c in table relation then ab is prime key so that all values can be determined by them and c and d fields are not prime fields. The table decomposition should eliminate the partial dependency. So the new table R(ab,d) and R(c,d) are in relation with primary key combination while design 2nd normal form Roll with Name and Roll with Course are designed to store all records.

Table no 5: Normalized from Table no 4

Roll	Name
105	Sila
105	Sila
105	Sila
106	Binod
107	Binod
108	Binod

Roll	Courses
105	Java
105	Oracle
105	Oracle
106	C++
106	Java
106	Computer

Similarly when table no 5 is re-structure on the basis of 2nd normal form to separate two tables are redraw with similar with table no 7 with relational table student info (Roll no, Name, Phone, State, Country, Age) and course info (Roll no, Course no, Course name).

Table no 6: Un- Normalized updated student database

Roll no	Name	Phone	State	Country	Age	Course no	Course name
1	Rupa	98467329	one	Nepal	21	3	C#
1	Rupa	98456432	one	Nepal	21	4	Oracle
2	Rupaa	98467392	two	Nepal	23	5	PHP
3	Sunita	98467329	one	Nepal	31	2	Python
4	Bimal	98467320	four	Nepal	22	4	Java

Table no 7: Normalized with 2NF of Table no 5

Rollno	Name	Phone	State	Country	Age
1	Rupa	98467329	one	Nepal	21
1	Rupa	98456432	one	Nepal	21
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22

Student info

Rollno	Courseno	Course name
1	3	C#
1	4	Oracle
2	5	PHP
3	2	Python
4	4	Java

course info

Foreign key column or combination of columns that are used to establish a relationship between data in two tables. The column used to create the primary key in one table are also used to create the foreign key constraint can be used to reference data in the same table or in another table. The foreign key does not have reference but primary key has reference either the same table or another table. The foreign key table may accept null but if it contain null the reference process is skipped.

Boyce Codd Normalization Form

BCNF is the fourth level of database administration that should be the 3rd normal form. The second focus of the normalization in the partial dependence allows α-> β to be the fields of the relational table in which α is the main candidate key or determining field of β attributes, in this case α is prime and β is a non-prime attribute. The second normal form cannot allow non-primary attributes to be depend on a part of the candidate key α. Similarly to become the 3rd normal form, the first and second normal form must be there and α-> β in this case both αβ must be a non-prime attribute so that non-primary attributes can find other non-primary attributes known as transitive dependency. To become BCNF, both the partial condition and the transitive dependence could not match. Suppose that the first β, which eventually discovers the attributes, indirectly forms the BCNF forms. If R is the highest normal form for every functional dependency, left hand side is non-trivial functional dependence is super-key. Suppose in relational table R (A,,B,C,D,E) then BC-> D, AC-> BE and B -> C are relation then AC become the candidate key in which A or B cannot be derived from any other attribute, so there is single candidate key (AC). The main attributes are part of the AC candidate key. Similarly, BDE are non-primary attributes, the relationship is first normal and does not allow multiple values or compound attributes. The BC-> D relation is in second normal form in which BC is not configured with the candidate key of AC and AC-> BE is also the candidate key B-> E is 2NFbut B is not the subset set of candidate key AC. The relation is not the third normal form because BC-> D and BC is super key or D is the main attribute. In B-> E, B is super key nor is the main attribute E. To become BCNF LHS, functional dependency must be super-key or RHS must be the main attribute. A relational schema R is in the normal form of Boyce-Codd, if and only if for each of its dependencies $X \rightarrow Y$, at least one of the following conditions is met:$X \rightarrow Y$ is a trivial functional dependency ($Y \subseteq X$) and X is a super key for the R schema.

FUTURE TRENDS IN THE DATABASES

Now is the urgent to look the future perspective of data science that present growing style of data management will meets future of databases design. Currently data scientist are starting to store in complex logic with separation of database itself; This feature allows a database to reside in more than one location and to be queried as a continuous unit which is called distributed or federated databases in data structure. While accessing data sets all intermediate servers run simultaneously in all locations.

This is possible due to the increase in network speed. (Berg, 2013) predicts the following trends in the database world. Although data scientist are changes in architecture, such as cloud computing and the need to manage large amounts of data. An example is Google Map Reduce. Data collections, not databases, are increasingly important for acquiring and connecting knowledge. Cloud, mobile and virtual applications are money changers. Data storage becomes increasingly complex which requires the need for specialized functions. It is also necessary to store image data, scanned data and complex medical data, such as gene sequences. Medical record physical data and transform it into a digital format are in use and must be processed by computers. Processing this data to derive models, such as DNA sequences that cause a particular disease, also requires more storage and processing of data. However, (Guy Harrison, 2018) Dell said that the end of a database valid for all dimensions because the single architecture cannot satisfy the request for availability, data volume, and data transaction for

all types of electronic devices are used in the modern world. Web server, web DBMS, memory management (Spark) Hadoop, RDBMS with oracle, Operational RDBMS, SQL SHAS Analytics HANA, and ERP in-house CRM are available for data management application in modern world.

Data variety and speed of data management in the modern world of the revolution of the Bluetooth industry, 3G Wi-Fi, high-quality red, GPS, headphones, speakers, monitors, monitors, tools communication, silent alarm, podcast for the bedroom are modern needs silently maintain through in cloud.it has been used intelligently to access a commutative human being. For the most part, Big Data's personal data is Google, which is maintained by Google File System (GFS) and the server's back-end. In the DIR, Hadoop Map Reduce Managed Stack is the best Big Data technology that uses very important domains in the modern world. The third trend is NoSQL, which preserves the collection of data in several services, the data network and the read pattern only for slaves. Amazon uses constant hashing, uses the family of data in columns where ODBMS and XML use the document database. In the same way, the infinite record uses the basic data records. In summary, the database has several users, as a key value, based on tables, document base and chart base. The fifth trend is the end of the disk, in which we replaced disk storage with analytics, Volt DB, Spark Oracle 12 and cloud computing. Data Analysis is the process of systematically applying statistical techniques to describe and illustrate, condense and recap, and evaluate date. An essential component of data integrity is the accurate and appropriate analysis of research findings.

Therefore, data-driven science, is an interdisciplinary field about scientific methods, processes, and systems, or knowledge or insights from data in various forms, either structured or unstructured, similar to data mining is based on functional dependency and normalization of all tabular records so that future generation would easily processed.

CONCLUSION

However, future of databases appears bright with possibilities when data scientist could design uniform API which could handle and analyzing data huge amount of data system. Electronic database and improving technology has made people all come to click or call away. Distributed transaction processing is becoming the norm for business planning in many countries. Data storing data base management shows variation in storage, size, and its architecture due to large organization needs. This can be made uniform in some extent from using functional dependency a d normalization techniques. This system is useful in data management context of data storage, large data in organized manner.

AUTHOR INFORMATION

Yagyanath Rimal is an assistant professor of Computer Science and Information Systems at Pokhara University,Nepal. He received his in Master of Science in Information Technology from Sikkim Manipal University Health Sciences with Information Systems Specialization he has wide area of interest of design web application, data base management and writing books.

REFERENCES

Berg, K. S. (2013). History of Database. International Journal of Management & Information Systems – First Quarter 2013, 3.

Ciupa, M. (2018). How the History of Data Gathering Lead to the Age of Big Data. FRY, J. (1976). Evolution of Database Management System. 2.

Gate, R. (2013). International Journal of Management & Information Systems Volume 17, Number 1. Kristi Berg,, 2.

Gates, P. (2017, 3). https://www.datavail.com/blog/4-top-trends-db-management/. Retrieved from https://www .datavail.com/blog/4-top-trends-db-management/: https://www.datavail.com/blog/4-top- trends-db-management/

Gray, J. (2017). Data Management: Past, Present, and Future . Microsoft Research IEEE Computer 29(10): 38-46 .

Guy Harrison. (2018). Dell Software Group. Retrieved from Dell and Quest – a brief history. Hayes, B. (2018). Analytics, Data Science, Machine Learning.

Lohr, S. (2016). New York Times.

Markarian, J. (2017). The Data Advantage. Retrieved from https://www.infoworld.com/article/3245644/technology-business/7-technology-trends-to-watch-in- 2018.html

McKendrick, J. (2017). Hot Data Technologies and Trends for 2018. Retrieved from Hot Data Technologies and Trends for 2018: http://www.dbta.com/Editorial/Trends-and-Applications/Hot-Data- Technologies-and-Trends-for-2018-122280.aspx

Peng, D. (2017). Retrieved from https://www.statista.com/statistics/274774/forecast-of-mobile-phone- users-worldwide/

Pinola, M. (2017). Data Storage Technologies of the Future. Retrieved from backblaze.com: https://www.backblaze.com/blog/data-storage-technologies-of-the-future/

Press, G. (2018, 3 7). Retrieved from https://www.forbes.com/sites/gilpress/2016/03/23/data- preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/2/#5d5ee48f1492 Sangeeta, G. a. (2016). DATA MINING TREND IN PAST, CURRENT AND FUTURE.

Shakir Khan, D. A. (2012). Data Mining For Securiy Purpose. (I. 2277, Ed.) INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH , 2.

U.M. Fayyad, G. P.-S. (1995). Advances in Data Mining and Knowledge Discovery,. Cambridge: MIT Press.

Whitehorn, M. (2017, 3). Get six expert tips for improving data warehouse performance. Learn how database engines, SSDs and MOLAP cubes can affect your data warehouse performance. Retrieved from http://searchdatamanagement.techtarget.com/answer/Six-tips-for-improving-data-warehouse- performance: http://searchdatamanagement.techtarget.com/answer/Six-tips-for-improving-data- warehouse-performance

Instructions for Authors

Essentials for Publishing in this Journal

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

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

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

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

Submission Process

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

Manuscript Guidelines

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

Title

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

Letterhead Title

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

Author's Name

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

Contact Details

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

Type of Articles

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

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

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

Language

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

Abstract and Summary

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

Keywords

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

Acknowledgements

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

Tables and Illustrations

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

Citation in the Text

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

Footnotes

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