Feature Extraction of Real-World Traces Using K Means Clustering In OppNet

Rahul Johari (rahuljohari@hotmail.com)
Guru Gobind Singh Indraprastha University https://orcid.org/0000-0002-7675-8550

Tanvi Gautam
Guru Gobind Singh Indraprastha University

Research

Keywords: Adhoc Network, Delay Tolerant Network, Intermittent Networks, Mobile Adhoc Network (MAN), Mobility Models, Opportunistic Networks (OppNet)

Posted Date: October 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-915827/v1

ARTICLE TYPE

Feature Extraction of real-world traces using K means clustering in OppNet

Rahul Johari*[1] | Tanvi Gautam[2]

[1]University School of Information and Communication Technology, GGS Indraprastha University, New Delhi, India
[2]University School of Information and Communication Technology, GGS Indraprastha University, New Delhi, India

Correspondence
*Rahul Johari, GGSIPU.
Email: rahul@ipu.ac.in

Present Address
Present address

Abstract

Natural calamities leave people helpless by arising several situations such as network breakdown, zero communication, intermittent connectivity, dynamic network topology. In such situation an application of dynamic and intermittent routing scheme is essential to make further communication possible during likewise scenarios. An application of TCP/IP becomes futile in mentioned circumstances as it best works for static nodes and pre-defined network topology wherein source and destination nodes are first establishing the communication link with each other. An alternative measure of such hitches is to encounter an application of DTN protocol which possess all characteristics to withstand in such scenarios such as; dynamic network topology, intermittent connectivity, frequent path breaks, store – carry – forward fashion. In this paper we did thorough investigation of forest fire dataset (Uttarakhand) after exploring its implementation in ONE with Epidemic, Prophet, Spray and Wait, HBPR, GAER respectively. An extensive and thorough investigation for real world traces implementation has been done with OppNet routing protocols against mobility models namely; Shortest path map – based, Random Direction, Random Walk, Random Waypoint, Cluster Movement respectively for network performance metrics namely packet delivery ratio, packet overhead ratio and average latency ratio respectively with the application of K means clustering machine learning algorithm. With the help of this analysis, we explore the real-world traces characteristics and study the areas on which

KEYWORDS:

Adhoc Network, Delay Tolerant Network, Intermittent Networks, Mobile Adhoc Network (MAN), Mobility Models, Opportunistic Networks (OppNet)

1 INTRODUCTION

Natural calamities such as cyclone, earthquakes, hurricanes, volcano eruption, flood etc. are nothing but a consequence of bad climate, global warming and several other reasons due to which several human activities has been affected such as communication gets crippled, network infrastructure breakdown, emotional and physical health problems, danger of death or physical harm, road destruction etc. One of the main functional paradigm is communication in such scenarios for making the necessity things possible to tackle with such situations like calling of an ambulance, making communication establishment feasible remotely, deployment of communication infrastructure, construction of network topology etc. hence under former mentioned situations wherein services such as WLL, CDMS, GSM,PSTN and so on has been disable or becomes non – functional, a network concept has been proposed by researchers called as Delay Tolerant Network (DTN).

DTN is nothing but a subclass of Wireless Area Network (WAN) which possess characteristic like; sparse network, store – carry – forward mechanism, dynamic network topology, mobile nodes, infrequent network contact time. situations likewise wherein nodes need to establish communication when frequent network breakdown has been occurred due to any reasons like natural calamities then DTN application has become invulnerable, then Opportunistic Network (OppNet) plays an important role. OppNet is an extension of DTN which possess all inherited characteristics of DTN in addition also; intermittent connectivity, store – carry – forward routing mechanism, applicable in disaster and emergency situations. But due to the following mentioned reasons certain times application of DTN becomes less tolerant as:

- Whenever there is a need of frequent network construction to establish communication path between source and destination as soon as they came into communication range such as called as an intermittent connection.
- During a routing if a node carrying a message packet in an intermittent connectivity due to mobile nature of node it becomes out of communication range with the communicating node, then application of DTN becomes failure.
- In an emergency situation there is a frequent need of communication establishment with the node which is in a communication range of another node then construction of dynamic network is required.
- Nodes need to store a message packet till it further comes into communication range with another node therefore application of store – carry – forward routing mechanism in essential resulting in an effective routing mechanism.

But in scenario when a node doesn't have sufficient space to store a message longer and withstand in an intermittent dynamic network connectivity application of DTN also becomes obsolete. Many researchers focused nowadays in OppNet because of node intermittent connectivity. In scenario where a node storage capacity of a message packet is less, scalability of a node in an environment where multiple participating nodes perform routing mechanism therein took a long time for connectivity and message routing. Security and high node mobility yet again a concern for connectivity of OppNet. Also, nodes establish an intermittent connectivity with each other as soon as node finds a node in a communication range with another node.

The main contribution of this work conducted are mentioned below:

i. **Features identification:** A machine learning algorithm has been used for studying the features of real-world mobility traces through study of network performance metrics such as packet delivery ratio, packet overhead ratio and average latency ratio respectively.

ii. **Model used for feature identification:** K means clustering based machine learning algorithm has been used which is a type of an unsupervised algorithm. It is a popular data analysis algorithm which has been used for making the best feature identification of the real-world traces.

iii. **Evaluation functions:** The feature identification of best machine learning model has been chosen on basis of network performance metrics.

Furthermore, this research paper has been formulated as follows:

- Section 1 gives an introduction of OppNet followed by the OppNet characteristics with inherited features of DTN and MANET (Mobile Adhoc Network).
- Section 2 literature survey of a paper explains all the work that has been proposed by researchers in the field of OppNet and implementation in a real – life dataset traces under the application of machine learning algorithms.
- Section 3, explains the experimental work that has been conducted of real- world mobility trace (Uttarakhand wild – life forest fire dataset) under mobility models namely; Shortest path map – based, Random Direction, Random Walk, Random Waypoint and Cluster Movement respectively against the existing routing protocols namely Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively in ONE simulator.
- Section 4 explains the observations and result analysis that has been presented for the three network performance metrics as: average packet delivery ratio, average latency ratio and average packet overhead ratio respectively.
- Section 5 presents the future work and conclusion section which presents all the features that has been extracted after application of K – means and on the basis of which we all explore the research areas on which respective OppNet routing protocol performance can be improved.

Method of Experimental analysis:

Following are the steps with the help of which we can do the implementation: -

Step1: Importing the real-world traces file into ONE simulator

Step2: Writing a program for converting the real-world traces file into (.wkt) format that are compatible within ONE simulator, if not in a format as compatible with ONE.

Step3: Extending each respective routing protocol file with the help of extending class concept used in java.

Step4: Exporting a perl script file "BonnMotion converter" for making a conversion of mobility traces with ONE compatible BonnMotion form.

Step5: Noting a reading of the result generated in report folder using "extending concept of java" for respective parameter.

Step6: Again, importing a file for "reinforcement learning, K – means, KNN (K nearest neighbors)" machine learning algorithms respectively.

Step 7: Noting a reading of the result generated in report folder using "extending concept of java" for respective parameter.

Step 8: Creating a graph and table for performing the comparative evaluation.

2 LITERATURS SURVEY

The main principle of DTN **[1]** is store – carry – forward fashion mechanism which makes the communication possible among intermittent networks which states that if a packet stores a message packet till it encounter a destination node which is in communication range of neighbouring node. In a network architecture former routing scheme is possible in the Bundle layer **[2]** of a network layer which is present between the application and network layer. An instance wherein node encounter another node which is a range of another node and it start forwarding a message packet till its deliver to the destination and continue to follow this routing scheme till the node is in communication range of another node, this principle is called as an intermittent connectivity which is nothing but an extension of DTN called as an OppNet **[3]**. Many researchers **[4] [5] [6] [7] [8] [9] [10]** conclude the following research areas wherein the performance of OppNet can be improve and optimized by applying few algorithms of machine learning, artificial intelligence, IoT:

- Scalability metric of a node **[11]**: "Scalability" can be defined as a factor imposed on the node extensibility that is node able to perform a multiple functions such as message forwarding, storage of a message packet, maintaining a routing table and so on without any hitches. The data dissemination needs to be taken place in an efficient way so as to make services available in a reliable and efficient form to all participants.
- High resource constraint of a node **[6]**: It can be observed that a node is able to store a message packet till it encounter another node which is in a communication range of another node to make an effective routing mechanism possible. Due to which a node requires high network resource constraints such as buffer usage, maintenance of a routing table, handling flow and congestion control in a network, power consumption and so on.
- Intermittent connectivity and store – carry – forward routing scheme **[7]**: Intermittent connectivity is nothing but a type of a sparse network wherein a node forms a network as soon it encounters another node which is in a communication

range of another network. A network that has been formed intermittently called as a dynamic network construction wherein a node has been formed with the participation of a mobile nodes and nodes signaling for a network range.

- Congestion level of a node **[8]**

It can be concluded that optimization of a network resources is an essential criterion for making the high usage of OppNet routing protocols in a real-life scenario. For making the routing mechanism efficient most important criteria in mind of a source node is packet delivery at the destination node.

In **[12]** Deepak et. al. proposed routing scheme kROp, which primarily works on the dynamically available network features for the selection of a node that is potentially good for packet routing over a network. This scheme incurs a good packet delivery ratio but somehow poor ratio in comparison with ProWait. Additionally, security issue and energy consumption are being high for this scheme which results in less optimized network. For future direction work more unsupervised algorithms can be applied to see the best changes incurred by the respective scheme.

In **[13]** Dhurandher et.al. proposed a work focused towards the network optimization scheme by determining the rate of success of packet delivery ratio of a node through the application of algorithms namely decision tree and neural networks. The proposed scheme MLProph which is a novel based routing scheme works on the principle of next hop selection after the delivery probability computation using the concept of machine learning and Prophet network performance parameters such as node speed, buffer size and contact history. The flaw of this scheme is that the implementation could not be taken place with real world mobility traces.

In **[14]** Carlos O. Rolim et.al. makes the use of predictive routing agents which works on the principle of determining the future values of nodes to make the routing efficient and optimizing the network resources. It basically works on the hypothesis information regarding the contextual information of a node which is possible with the help of a multi – staging system wherein an agent is assigned to the whole network with a responsibility to determine context values of a node that consist of node future values and context information of a node.

In **[15]** Shubham et.al. proposed an intelligent routing scheme called “intelligent water drop algorithm” works on an optimization routing scheme in which dynamic next hop selection done. The node weights calculation done with the help of IWD algorithm wherein a node buffer space, number of message packets delivered, energy levels by applying a transitive function for routing of a message packet among dynamic network. This scheme incurs with the flaw of that node needs to recharge after a while to improve the packet delivery ratio and for traversing a routing path.

In **[16]** works on Ant algorithm which is a heuristic probabilistic routing scheme works on the principle of finding a good and short routing path among source and destination node for message packet delivery. In this proposed scheme by researchers, it mainly focusses in making the network optimize so as to minimal the network resource consumption. The advantage of this scheme is that it incurs less resource consumption ad gives high packet delivery ratio wherein main flaw of scheme is less packet delivery ratio and high resource consumption.

In **[17]** Xulin et.al. proposed a routing scheme in which the link prediction has been done on the basis of network topology, available network resources and node information respectively and framework called as RNN – LP approach. It is a neural network-based network link approach in which a vector is obtained that comprises of node present and past contextual information using time series analysis. RNN – LP provides better accuracy in terms of delivery and access ratio. In Table 1 we present the tabular study of the literature survey that we had conducted.

Table 1. Literature survey

Paper title	Objective	Methodology	Inference parameters	Pros	Cons
kROp k means routing based protocol for Opportunistic networks **[12]**	Identification of high potential and good forwarder nodes	K means clustering algorithm	Hop count, dropped messages, delivery probability, overhead ratio	Good message delivery probability and number of hop counts are less	Inclusion of more node features. Energy consumption and security issue not taken into account. Poor latency ratio.

A machine learning based routing protocol for OppNets **[13]**	Determining the successful delivery probability ratio on basis of P_L	Decision tree Neural network	Number of dropped/successful message delivery ratio, overhead ratio, hop count, buffer space usage	Good performance in comparison with Prophet	No simulation in real world traces. Use of ML classifiers.
Towards predictive routing agents in OppNets **[14]**	Multi staging agent system	Use of node contextual information for routing path prediction.	Message delivery Packet overhead ratio	Reasonable performance of message delivery and overhead ratio	Strategy can be proposed for reducing the time. Nodes needs to get recharge timely.
Neural network-based routing protocol for OppNet using IWD **[15]**	Intelligent routing decision using high robustness of network (IWDNN)	Intelligent water drop optimization	Delivery ratio, average latency, overhead ratio, drop ratio average buffer time vs message generation interval	Superiority in respect to message delivery ratio	Verification of network robustness. Learning rate and neural network momentum parameters left to be study.
Ant Router: An efficient routing protocol for social OppNets **[16]**	To increase routing efficiency by traversing a shortest routing path availability	Ant colony optimization algorithm	Number of nodes, buffer size, message size, message generation interval	Possess practical applications. Significant improvement over network performance metrics.	Message must reach to destination through optimal routing path only. Security and Privacy of node not done.
Link prediction approach for OppNets based on neural networks **[17]**	Estimation of future network link availability	RNN – LP algorithm (Recurrent neural network link prediction) algorithm	90% high accuracy and stability through random sampling of traces.	Achieves better network link prediction Better stability and accuracy	Dynamic network topology gains high overhead ratio.

3 PROBLEM STATEMENT

Former routing protocols that have been proposed in OppNets are basically limited to message forwarding only not to make the routing path optimal, utilizing less network resource consumption and further improvement. In the proposed research work we are basically focusing on the objective of feature extraction of the real-world mobility traces on the basis of its implementation that has been done under the K means clustering algorithm. The implementation has been presented for the OppNet routing protocols; Epidemic, Prophet, Spray and Wait, HBPR and GAER against mobility models namely; Shortest path map based, Random Direction, Random Walk, Random Waypoint and Cluster Movement respectively in terms of packet delivery ratio, access ratio and overhead ratio respectively. Afterwards, a comparative study has been done for the former objective implementation and we listed all numerous solutions on the loophole areas wherein improvement can be done to make the optimal routing and less network resource consumption.

3.1. About the dataset: Uttarakhand has been ranked as a 2nd state of India after Madhya Pradesh in the terms of forest fire incidents. In the year 2020, a forest fire incident has been occurred in Uttarakhand in October 1, 2020 and it will continue to April 4, 2021 causing a high damage of 1,297.43 hectares of forest, destruction of flora and fauna, death of wild life fires and loss of land fertility respectively. Latest study **[18]** shows that Uttarakhand regions four national parks namely Jim Corbett National Park, Govind wildlife sanctuary, Nandadevi forest division and Rajaji tiger reserve are highly prone areas for occurring of forest fires. According to National Interagency Fire Centre report Uttarakhand has 45.32% of its geographical area under the forest cover that includes snowbound landscapes and alpine pastures as forest total area is as high as 71.05%. Within 2005 to 2015, it has been witnessed around 601 incidents of forest fires approximately that us 66.39% total damage of wildlife damage (consist of animals and plants both). In our research work following former mentioned objectives in which we need to do an investigation study of OppNet existing routing protocols implementation in presence of real-world traces **[19]** of Uttarakhand wildlife sanctuaries and forest areas. Figure 1, showing the image trace of wildlife sanctuaries and forest regions covered in Uttarakhand.

3.2. Why K means clustering algorithm of machine learning: The purpose of using the existing machine learning algorithm in the proposed research work are as follows:

- Scrutinization of data: In OppNets it has been observed that the hidden part of datasets and identification of patterns are not clearly visible with OppNet routing algorithm. Scrutinization can be defined as "Analysis of an unknown data so that exploration of areas which are most likely to be group in terms of node movement, node routing path and node mobility.
- Traffic prediction and classification: In OppNet as network possess intermittent connections during routing hence it becomes difficult to lookout for the traffic pattern and classification of areas on the basis of participating traffic. With the application and implementation of machine learning algorithms **[20]** it is a necessity to check the traffic pattern on basis of existing machine learning algorithms. '
- Congestion control: Like Epidemic **[21]** which is a flooding-based routing protocol works on the flooding mechanism by routing a message packet to all participating nodes, it ensures 100% message delivery ratio but incurs high amount of resource consumption. So as to make the optimal resource utilization among all participating nodes implementation of machine learning algorithms becomes mandatory.
- Security and Fault management: Security and fault management of data during routing in OppNet is a difficult task as nodes forwarding a message packet to neighbouring nodes only for the purpose of forwarding a message packet till destination node. But detection of a loopholes/malware/bottleneck in traffic is a tedious task. Therefore, application of machine learning plays a vital role by applying certain encryption algorithms namely XGBoost or to reduce the depth of the data by applying logistic regression model on the dataset.

Figure 1. Uttarakhand wildlife sanctuaries and forest regions

4 SIMULATION

In our proposed research work the implementation of work has been done in ONE **[22]** simulator which is capable of performing the functions such as; generating node movement using numerous inbuilt available mobility models, routing of message packet among several neighbouring nodes, use of graphical user interface feature (gui) possess functionality of viewing route path, message packets movement, node structure and so on. ONE also has one distinguish feature an importing a real-world trace file and generating a report on the basis of its implementation using respective routing protocol. As our proposed work implementation has been subdivided into two main tasks; primarily an implementation of a real-world traces using Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively. Secondly, we did a comparative evaluation in which we compared the results of real-world traces with the existing machine learning algorithms namely K – means clustering algorithm and on the basis of this evaluation we evaluated areas in which we can improved the factors so as to optimize the routing path and network resources in terms of packet delivery ratio, access ratio and packet overhead ratio respectively.

4.1. Implementation of real-world mobility traces in ONE using routing protocols

In Section 4.1. we are implementing the real-world traces using existing routing protocols namely Epidemic, Prophet, Spray and Wait, HBPR and GAER. Afte conducting the experimental analysis of this implementation we come out an observation in terms

of packet delivery ratio, access ratio and packet overhead ratio respectively. On the basis of which we will conduct a comparison study on basis of former mentioned network features such as packet delivery ratio, packet overhead ratio and average latency ratio respectively. During the implementation of the work, we keep the simulation parameters constants such as number of nodes are those which are participating in the simulation process “n=200”, message ttl is the amount of time required to cache a queue and keeping it constant at 300ms. An interface is a media through all participating nodes communicate with each other such as “Bluetooth” so as to make possible the message forwarding mechanism possible among them. Event is a java class which needs to be extended to the chosen routing protocol, here it has been selected as “External Event” file is used for making the implementation possible for real world traces only. Report is a java class folder which needs to be imported with the help of extending concept of java and it basically comprises of all the status report about the network performance metrics such as number of hop counts. The size of message packet has been defined as 50KB – 1MB which is a small segment/whole message segment that has been forwarded over the network by a source to its destination. The traffic type that has been selected is “pedestrian” only. Number of messages dropped, number of messages generated, number of messages aborted, number of messages delivered, number of nodes connected and so on. Movement/Mobility models that have been chosen here are: Random Walk, Random Waypoint, Shortest path map based, Map based movement and Cluster movement model respectively. Router files that have been imported for router class are Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively.

Table 2. Default settings of ONE simulator

Default Setting for ONE	
Simulation connections	FALSE
Type of traffic	Only pedestrians
Number of host group	1 (n=200)
Msg ttl	300
Buffer Size	2000M
Message packet size	50KB - 1MB
Interface	Simple Broadcast
Type	Bluetooth
Group speed	0.0,0.0
Interface transmit range	10m
Group.okMaps	1
Event.class	External Event
Report.report	MessageStatsReport
Map based movement	"path to be given"
Group.movementModel	Random walk
	random waypoint
	shortest path map based
	map based movement
	cluster movement
Group. Router	Epidemic
	Spray & Wait
	Prophet
	HBPR
	GAER

5 RESULT AND DISCUSSION SECTION

In this Section we are elaborating the result that has been generated after conducting the simulation of mentioned real world traces in ONE simulator using K means algorithms. The simulation results obtained with the application of routing protocols namely Epidemic, Prophet, Spray and Wait, HBPR and GAER against the mobility models namely shortest path – map based, Random Walk, Random Direction, Random Waypoint and Cluster Movement respectively. The network parameters whose analysis has been done here are; packet delivery ratio, packet overhead ratio and average latency ratio respectively. In Table we are also showing the error rate (%) in terms of bottleneck probability for conducting the result analysis. We also calculated the ratio of all network parametric using the formula mentioned in Equation 1, 2 and 3 respectively.

Packet delivery ratio can be defined as transmitting ratio of total number of message packets delivered to the destination to the total number of message packets sent from source to destination node.

$$PDR = \frac{Number\ of\ packets\ delivered}{Total\ number\ of\ packets} \quad (1)$$

Packet overhead ratio is the ratio of total number of bytes generated/consumed by a sender for passing a message packet over a secured routing path or the total number of bytes transferred among sender and receiver over a network. Also representing the transmission cost of a network routing. It also reflects the amount of resource consumption done during a message packet forwarding.

$$POR = \frac{RM - DM}{DM} \quad (2)$$

Average latency ratio can be defined as total RTT taken by a message packet during a routing that is "a total amount of time taken by a message packet for routing a message packet over a network configuration so as to deliver it to the destination and coming back to its destination." The good ratio considered an optimal if it lies between 20ms – 40ms range.

$$\Sigma_{i=1}^{n} = \frac{Time\ at\ which\ message\ recieved - Time\ at\ which\ message\ created}{Total\ number\ of\ message\ recevied} \quad (3)$$

In Table 3, represents the results generated on the basis of mobility model namely shortest path map-based movement model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively. In Figure 2, graphical comparative study has been shown separately on the basis of routing protocols separately not in a cumulative form.

Table 3. Analysis of K – means using OppNet under Shortest path map-based

S.No.	Mobility model	Routing protocol	Packet delivery	Packet overhead	Average latency	Packet delivery ratio	Packet overhead ratio	Average latency ratio
1.	Shortest path map – based movement	Epidemic	194	190	141	97%	95%	70.%
2.		Prophet	183	181	114	91.5%	90.5%	57%
3.		Spray and wait	188	174	141	94%	87%	70.5%
4.		HBPR	164	94	74	82%	37%	37%
5.		GAER	172	98	58	86%	49%	29%

The following observations can be concluded from Table 3, that: -

- The packet delivery ratio of Epidemic is highest among all routing protocols which is 97% and lowest is of HBPR which is 82%. There is a lag of 15% for performance of HBPR in comparison with Epidemic. K means always choose the smallest centroid which allocated to all existing clusters as much as possible the lowest value has been assigned to the centroid. From the result analysis that is obtained after performing a simulation it can be seen that for shortest path map-based mobility model chosen centroid value for Epidemic is small as much as possible during simulation as Epidemic works on the principle of flooding mechanism (that is flooding a message packet to all neighbouring nodes). Performance of Spray and Wait lags with Epidemic by 3% whereas for Prophet by 5.5%. The order for packet delivery ratio is given as follows:

$$Epidemic > Spray\ and\ Wait > Prophet > GAER > HBPR$$

- During analysis of packet overhead ratio for K means under shortest path map – based mobility model performance of Epidemic again is on peak as in comparison with other routing protocols by 4.5%, 8%, 58%, 46% for Prophet, Spray and Wait, GAER, HBPR respectively. It can be concluded from above metrics that performance of HBPR is best in comparison with Epidemic, Prophet, Spray and Wait and GAER by 58%, 53.5%,50% and 46% respectively. Overhead ratio also denotes the transmission cost of network construction and maintenance, from analysis it can be seen that performance of HBPR is good in comparison with Epidemic by 58%. It also shows that overhead ratio of Epidemic is high which shows that for maintaining the network it is not a good option. The increasing order of packet overhead ratio has been shown below:

$$HBPR > GAER > Spray\ and\ Wait > Prophet > Epidemic$$

For average latency ratio the transmission rate in shortest path map – based mobility model good performance index rate is given by GAER which is a genetic concept-based routing protocol of OppNet. It can also be seen that overhead ratio of GAER is average in comparison with HBPR by 8% only. For HBPR it can be seen that performance rate is constant for both overhead ratio and average latency ratio but in GAER it can be seen that its overhead ratio ahead by 20% in comparison with its average latency ratio. The primarily reason for this difference is due to the mobility model that has been chosen here which always prefer to choose a shortest path on the basis of node movement in a network.

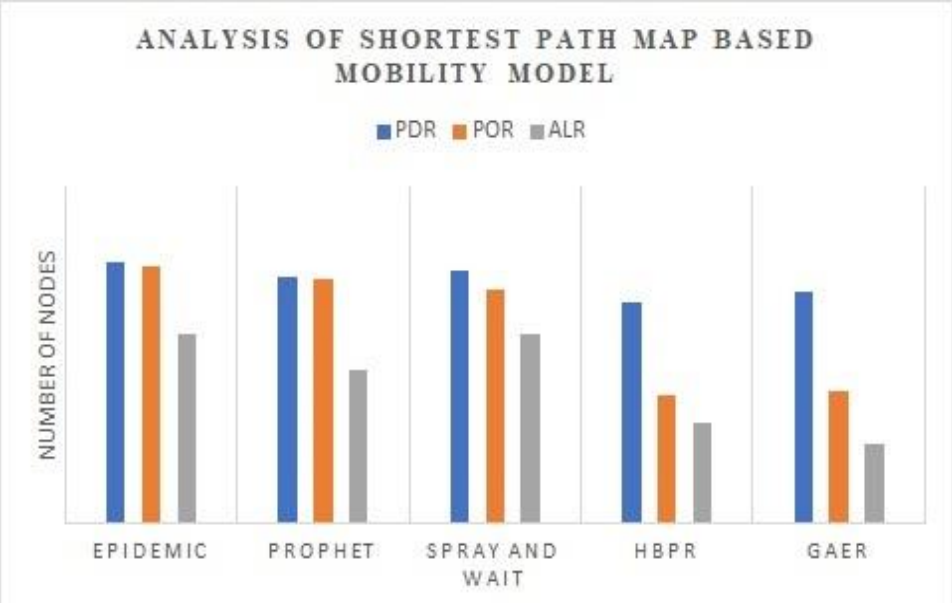

Figure 2. Analysis of Shortest path map – based

In Table 4, represents the results generated on the basis of mobility model namely Random Walk mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively. In Figure 3, graphical comparative study has been shown separately on the basis of routing protocols separately not in a cumulative form.

Table 4. Analysis of K – means using OppNet under Random Walk

S.No.	Mobility model	Routing protocol	Packet delivery	Packet overhead	Average latency	Packet delivery ratio	Packet overhead ratio	Average latency ratio
1.	Random Walk	Epidemic	194	190	184	97%	95%	92%
2.		Prophet	179	182	190	89.5%	91%	95%
3.		Spray and wait	65	84	41	32.5%	42%	20.5%
4.		HBPR	77	98	58	38.5%	49%	29%
5.		GAER	178	98	133	89%	49%	66.5%

The following observations can be concluded from Table 4, that: -

- In packet delivery ratio it can be seen that again for Random Walk performance of Epidemic is again 97% as in shortest path map – based mobility model whereas Spray and Wait gives a worse performance. The performance of Prophet and GAER is almost same and differ only by 0.5% whereas Epidemic performance ahead by 7.5% in comparison with them. The performance of HBPR is differ by 6% in comparison with spray and wait routing protocol. The increasing order for performance of packet delivery ratio is given by:

$$Epidemic > Prophet > GAER > HBPR > Spray\ and\ Wait$$

- For packet overhead ratio analysis, performance of spray and wait is best in comparison with Epidemic whereas HBPR and GAER performance is almost same 49%. Epidemic performance is worse in comparison with HBPR and GAER by 46% whereas Epidemic and Prophet performance is almost same and worse with a difference of 4%. Spray and wait performance are best in comparison with HBPR and GAER by 7%. The increasing order of packet overhead ratio is given by:

$$Spray\ and\ Wait > HBPR/GAER > Prophet > Epidemic$$

- Average latency ratio performance is best for Spray and Wait which is also same for packet overhead ratio that is 20.5% only. Similarly, performance of HBPR is average in comparison with spray and wait like packet overhead ratio which is 29%. The difference of HBPR is 8.5% in comparison with spray and wait which is nearly same like in packet overhead ratio. Here, performance of Prophet strike in comparison with Epidemic by a difference of 4% while for GAER it is 66.5% with a difference of 23.5%. The increasing order of average latency ratio is given by:

$$Spray\ and\ Wait > HBPR > GAER > Epidemic > Prophet$$

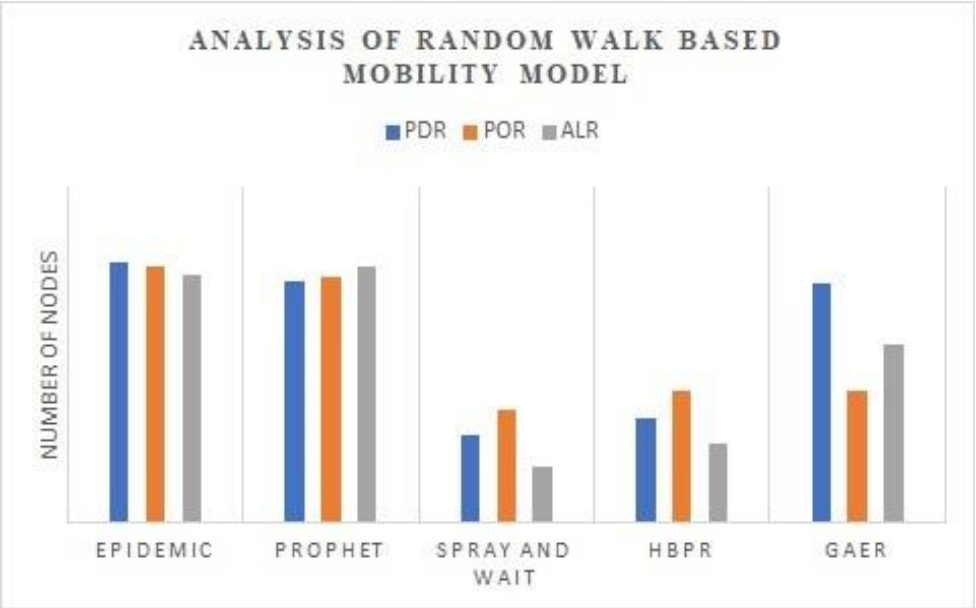

Figure 3. Analysis using Random Walk

In Table 5, represents the results generated on the basis of mobility model namely Random Walk mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively. In Figure 4, graphical comparative study has been shown separately on the basis of routing protocols separately not in a cumulative form.

- Using packet delivery ratio metric performance of Prophet is improved and best in comparison with Epidemic by 0.5% for Random Direction mobility model. Also, performance of GAER also improved in comparison with Epidemic by 2% only. The worse performance has been given by spray and wait and HBPR in comparison with Prophet and GAER by {57%,51%}, {56.5%,50.5%} respectively. The increasing order for packet delivery ratio is given by:

$$Prophet > GAER > Epidemic > HBPR > Spray\ and\ Wait$$

Table 5. Analysis of K – means using OppNet under Random Direction

S.No.	Mobility model	Routing protocol	Packet delivery	Packet overhead	Average latency	Packet delivery ratio	Packet overhead ratio	Average latency ratio
1.	Random Direction	Epidemic	198	114	103	99%	57%	51.5%
2.		Prophet	188	172	192	94%	86%	96%
3.		Spray and wait	114	148	104	57%	74%	52%
4.		HBPR	104	98	85	52%	49%	42.5%
5.		GAER	48	61	55	24%	30.5%	27.5%

- The packet overhead ratio not only a network performance metric which is used for performance assessment of network but also a cost metric. GAER performance is best in comparison with Prophet and Spray and Wait by a difference of 12%, 26.5% respectively. Worse performance is given by Prophet in comparison with GAER by 55.5%. The increasing order of packet overhead ratio is given by:

$$GAER > HBPR > Epidemic > Spray\ and\ Wait > Prophet$$

- For average latency ratio, worse performance has been given by Prophet which is same as in Random Walk and best is given by GAER with a difference of 24% only. The performance of Epidemic and Spray and Wait is almost same with a difference of 0.5% only. The performance ratio for average latency network metric is almost same for HBPR, Spray and Wait, Epidemic. The increasing order of average latency ratio is given by:

$$GAER > HBPR > Epidemic > Spray\ and\ Wait > Prophet$$

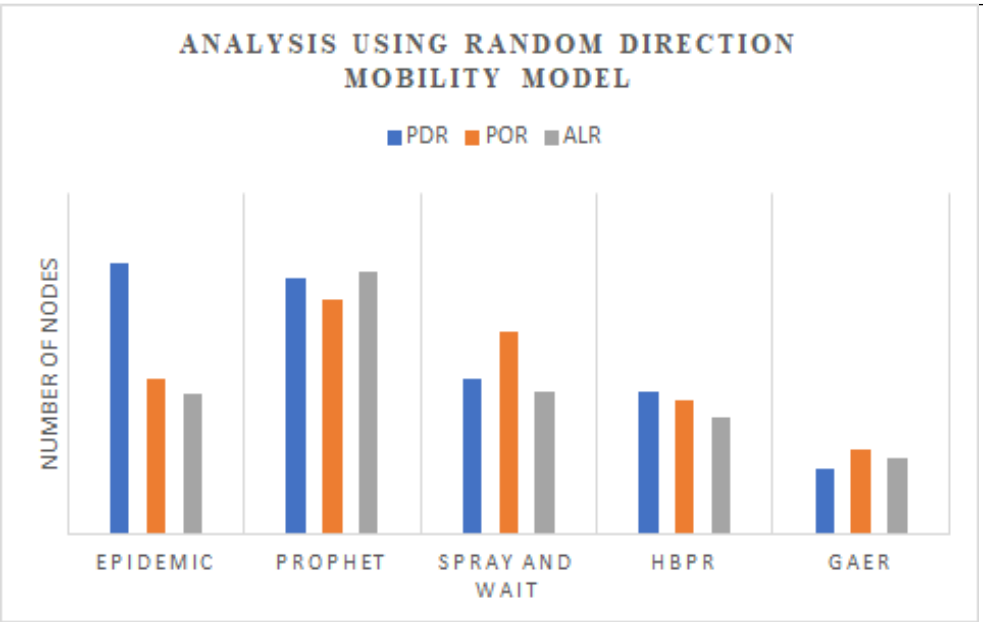

Figure 4. Analysis using Random Direction

In Table 6, represents the results generated on the basis of mobility model namely Random Waypoint mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively. In Figure 5, graphical comparative study has been shown separately on the basis of routing protocols separately not in a cumulative form.

Table 6. Analysis of K – means using OppNet under Random Waypoint

S.No.	Mobility model	Routing protocol	Packet delivery	Packet overhead	Average latency	Packet delivery ratio	Packet overhead ratio	Average latency ratio
1.	Random Waypoint	Epidemic	149	109	114	74.5%	54.5%	57%
2.		Prophet	141	118	138	70.5%	59%	69%
3.		Spray and wait	104	147	89	52%	73.5%	44.5%
4.		HBPR	174	42	154	87%	21%	77%
5.		GAER	175	65	14	87.5%	32.5%	7%

- In packet delivery ratio assessment best performance has been given by GAER for random waypoint mobility model which is 87.5% while worse performance has been given by spray and wait which is 52%. The performance of HBPR lags only by 1% in comparison with GAER while Prophet performance is 70.5% in comparison with Epidemic that lags by 4.5% only. The increasing order of packet delivery ratio is given by:

$$GAER > HBPR > Epidemic > Prophet > Spray\ and\ Wait$$

- For packet overhead ratio worse performance has been given by Spray and Wait which is 73.5% while for packet delivery ratio worse performance has been given by spray and wait too. Similarly, for HBPR best performance is presented while for packet delivery ratio it is just lags only by 1%. Also, for GAER good overhead ratio has been given in comparison with HBPR which has been just lags only 8.5%. The increasing order for overhead ratio has been given by:

$$HBPR > GAER > Epidemic > Prophet > Spray\ and\ Wait$$

- The average latency ratio best performance has been given by GAER which is also given same for packet delivery ratio but in case of packet overhead ratio it's just lags by 8.5% only. HBPR gives the worse performance which is 77% but for overhead ratio it gives optimal performance but in this case its value just shoots high at 77%. Similarly for Prophet worse performance is given only with a difference of 8% in comparison with HBPR while for Epidemic and Spray and Wait almost same performance has been given with a difference of 12.5% only.

$$GAER > Spray\ and\ Wait > Epidemic > Prophet > HBPR$$

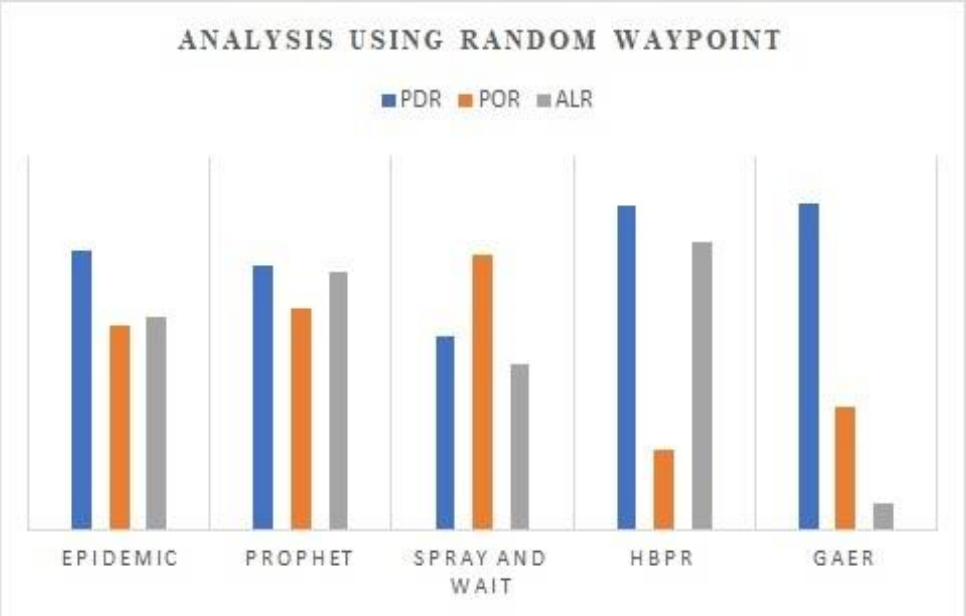

Figure 5. Analysis using Random Waypoint

In Table 7, represents the results generated on the basis of mobility model namely Cluster Movement mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively. In Figure 6, graphical comparative study has been shown separately on the basis of routing protocols separately not in a cumulative form.

Table 7. Analysis of K – means using OppNet under Cluster Movement

S.No.	Mobility model	Routing protocol	Packet delivery	Packet overhead	Average latency	Packet delivery ratio	Packet overhead ratio	Average latency ratio
1.	Cluster Movement	Epidemic	149	109	114	74.5%	54.5%	57%
2.		Prophet	141	118	138	70.5%	59%	69%
3.		Spray and wait	104	147	89	52%	73.5%	44.5%
4.		HBPR	174	42	154	87%	21%	77%
5.		GAER	175	65	14	87.5%	32.5%	7%

- In packet delivery ratio, best performance has been given by GAER which is 87.5% while worse performance ratio is given by Spray and Wait which is 52%. The performance ratio of HBPR nearly same to GAER only lags by 0.5% while performance ratio of Epidemic and Prophet is only difference with 4% with 74.5% and 70.5% respectively. The worse performance has been given by Spray and wait which is 52% only. The increasing order for packet delivery ratio performance is given by:

$$GAER > HBPR > Epidemic > Prophet > Spray\ and\ Wait$$

- The packet overhead ratio is basically denoting the performance assessment for optimal network by conducting an analysis in terms of transmission cost, network parameters requirement. Here for cluster movement mobility model network routing protocol which gives best performance is HBPR while Spray and wait again gives worse performance for the overhead ratio. Performance ratio for GAER is nearly same for HBPR which is 32.5% and lags by 8.5% in comparison with it. The increasing order for performance assessment is given by:

$$HBPR > GAER > Epidemic > Prophet > Spray\ and\ Wait$$

- For average latency ratio, worse performance is given by Prophet which basically works on the delivery predictability value while best performance is given by GAER again at 7% which is same in case of Random Waypoint. The performance ratio of random waypoint differs only by few percent in comparison with the Cluster Movement mobility model because both works on the concept of clusters but random waypoint forms the cluster after the performance assessment. The increasing order for average latency value is given by:

$$GAER > Spray\ and\ Wait > Epidemic > Prophet > HBPR$$

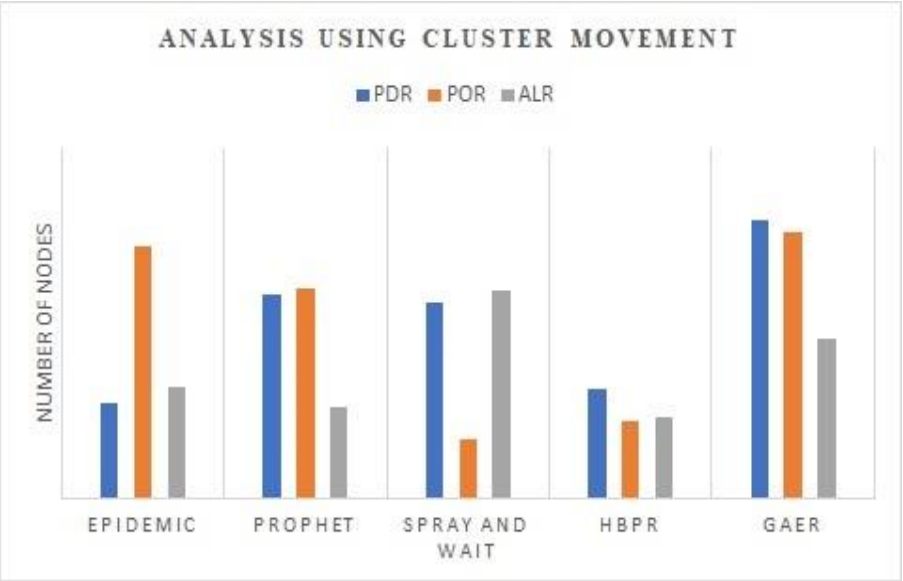

Figure 6. Analysis using Cluster Movement

6 FUTURE WORK AND CONCLUSION

In this research work we basically focus on the feature extraction of real-world mobility trace (Uttarakhand Forest fire) with the help of k means clustering machine learning algorithm that primarily focus on studying the highest prone areas of Uttarakhand regions wherein there is a high probability of forest fire occurring on the basis of performance assessment in terms of metrics such as packet delivery ratio, overhead ratio and average latency ratio respectively. It can be seen that performance ratio of GAER and HBPR is good for the mobility models namely Random waypoint and cluster movement individually while for random direction and random walk performance ratio of Epidemic and Prophet is good. For future work we can see the directions for working in the development of a routing protocol on the basis of features extracted which can be easily seen in Table 8, can propose a good and optimal routing protocol.

Table 8. Conclusion from K means analysis on Uttarakhand Forest fire traces

Mobility model	Routing protocol	Packet delivery ratio	Packet overhead ratio	Average latency value
Shortest path map based	Epidemic	Outstanding	Worse	Nominal
	Prophet	Best	Nominal	Average
	Spray and Wait	Average	Average	Worse
	HBPR	Best	Outstanding	Best
	GAER	Nominal	Best	Outstanding
Random Walk	Epidemic	Outstanding	Worse	Best
	Prophet	Best	Average	Worse
	Spray and Wait	Worse	Outstanding	Outstanding
	HBPR	Nominal	Best	Good
	GAER	Average	Best	Nominal
Random Direction	Epidemic	Outstanding	Nominal	Nominal
	Prophet	Best	Worse	Worse
	Spray and Wait	Average	Average	Average
	HBPR	Nominal	Best	Best
	GAER	Worse	Outstanding	Outstanding

Random Waypoint	Epidemic	Average	Nominal	Best
	Prophet	Nominal	Average	Average
	Spray and Wait	Worse	Worse	Nominal
	HBPR	Best	Outstanding	Worse
	GAER	Outstanding	Best	Outstanding
Cluster Movement	Epidemic	Average	Average	Nominal
	Prophet	Nominal	Nominal	Average
	Spray and Wait	Worse	Worse	Best
	HBPR	Best	Outstanding	Worse
	GAER	Outstanding	Best	Outstanding

Declaration

Availability of Data and Materials

Data sharing not applicable to this article as no datasets were generated during the current study.

Competing interest

The authors declare that they have no competing interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, not – for – profit sectors.

Author's Contributions

RJ's contribution is to provide the real-world mobility traces files for conducting an implementation and simulation in ONE, guiding in the paper organization, revise the paper. TG's contribution is to write a paper, conducting a simulation and performance analysis.

Acknowledgment

This work was supported by GGS Indraprastha University and I would like to thank Mr. Rahul Johari for his support to conducted this work simulation.

REFERENCES

[1] Hom, Jacqueline, Lauren Good, and Shuhui Yang. "A survey of social-based routing protocols in delay tolerant networks." In *2017 International conference on computing, networking and communications (ICNC)*, pp. 788-792. IEEE, 2017.

[2] Gautam, Tanvi, and Amita Dev. "Opportunistic network routing protocols: challenges, implementation and evaluation." In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 100-106. IEEE, 2019.

[3] Dede, Jens, Anna Förster, Enrique Hernández-Orallo, Jorge Herrera-Tapia, Koojana Kuladinithi, Vishnupriya Kuppusamy, Pietro Manzoni, Anas bin Muslim, Asanga Udugama, and Zeynep Vatandas. "Simulating opportunistic networks: Survey and future directions." *IEEE Communications Surveys & Tutorials* 20, no. 2 (2017): 1547-1573.

[4] Garg, Puneet, Ashutosh Dixit, and Preeti Sethi. "Link Prediction Techniques for Opportunistic Networks using Machine Learning." *Available at SSRN 3842849* (2021).

[5] Vashishth, Vidushi, Anshuman Chhabra, and Deepak Kumar Sharma. "A machine learning approach using classifier cascades for optimal routing in opportunistic internet of things networks." In *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1-9. IEEE, 2019.

[6] Gao, Yang, Jun Tao, Yifan Xu, Zuyan Wang, Weice Sun, and Guang Cheng. "CEBD: Contact Evidence-Driven Blackhole Detection Based on Machine Learning in OppNets." *IEEE Transactions on Computational Social Systems* (2021).

[7] Abbas, Salem, and Khaleel Ahmad. "Opportunistic Routing Protocols Based on Fuzzy Logic: Present and Future Directions." In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 1055-1060. IEEE, 2021.

[8] Cuka, Miralda, Donald Elmazi, Ryoichiro Obukata, Kesuke Ozera, Tetsuya Oda, and Leonard Barolli. "An integrated intelligent system for IoT device selection and placement in opportunistic networks using fuzzy logic and genetic algorithm." In *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 201-207. IEEE, 2017.

[9] Borah, Satya J., Sanjay Kumar Dhurandher, Isaac Woungang, and Vinesh Kumar. "A game theoretic context-based routing

protocol for opportunistic networks in an IoT scenario." *Computer Networks* 129 (2017): 572-584.

[10] Cuka, Miralda, Donald Elmazi, Ryoichiro Obukata, Kesuke Ozera, Tetsuya Oda, and Leonard Barolli. "An integrated intelligent system for IoT device selection and placement in opportunistic networks using fuzzy logic and genetic algorithm." In *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 201-207. IEEE, 2017.

[11] Valerio, Lorenzo, Andrea Passarella, Marco Conti, and Elena Pagani. "Scalable data dissemination in opportunistic networks through cognitive methods." *Pervasive and Mobile Computing* 16 (2015): 115-135.

[12] Sharma, Deepak Kumar, Sanjay Kumar Dhurandher, Divyansh Agarwal, and Kunal Arora. "kROp: k-Means clustering based routing protocol for opportunistic networks." *Journal of Ambient Intelligence and Humanized Computing* 10, no. 4 (2019): 1289-1306.

[13] Sharma, Deepak K., Sanjay K. Dhurandher, Isaac Woungang, Rohit K. Srivastava, Anhad Mohananey, and Joel JPC Rodrigues. "A machine learning-based protocol for efficient routing in opportunistic networks." *IEEE Systems Journal* 12, no. 3 (2016): 2207-2213.

[14] Rolim, Carlos O., Anubis G. Rossetto, V. R. Leithardt, Guilherme A. Borges, T. F. Dos Santos, Adriano M. Souza, and C. F. Geyer. "Towards predictive routing agents in opportunistic networks." In *Fifth International Workshop on Collaborative Agents Research & Development, CARE for Intelligent Mobile Services-Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Springer-Verlag Berlin Heidelberg Paris, France, 2014.

[15] Kumaram, Shubham, Shefali Srivastava, and Deepak Kumar Sharma. "Neural network-based routing protocol for opportunistic networks with intelligent water drop optimization." *International Journal of Communication Systems* 33, no. 8 (2020): e4368.

[16] Sharma, Deepak Kr, Swati Singh, Vishakha Gautam, Shubham Kumaram, Mehul Sharma, and Shrid Pant. "Ant Router: An efficient routing protocol for social opportunistic networks using ant routing." *IET Networks* 9, no. 2 (2019): 83-93.

[17] Cai, Xulin, Jian Shu, and Manar Al-Kali. "Link prediction approach for opportunistic networks based on recurrent neural network." *IEEE Access* 7 (2018): 2017-2025.

[18] Kumar, Sunil, Gopal Das Bairagi, and Amit Kumar. "Identifying triggers for forest fire and assessing fire susceptibility of forests in Indian western Himalaya using geospatial techniques." *Natural Hazards* 78, no. 1 (2015): 203-217.

[19] Alireza Shamsoshoara, Fatemeh Afghah, Abolfazl Razi, Liming Zheng, Peter Fulé, Erik Blasch. "The FLAME dataset: Aerial Imagery Pile burn detection using drones (UAVs)." doi: 10.21227/qad6-r683

[20] Das, Kajaree, and Rabi Narayan Behera. "A survey on machine learning: concept, algorithms and applications." *International Journal of Innovative Research in Computer and Communication Engineering* 5, no. 2 (2017): 1301-1309.

[21] Vahdat, Amin, and David Becker. "Epidemic routing for partially connected ad hoc networks." (2000): 2019.

S.No.	Figure No.	Figure title	Figure description (50words)
1.	Figure 1	Uttarakhand wildlife sanctuaries and forest regions	Figure 1, representing an image of the Uttarakhand Forest area which is the high prone areas of Uttarakhand Forest areas with high probability of forest fires disaster occurrence.
2.	Figure 2	Analysis of Shortest path map – based	Figure 2, represent the results generated on the basis of mobility model namely shortest path map-based movement model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively
3.	Figure 3	Analysis using Random Walk	Figure 3, represent the results generated on the basis of mobility model namely Random Walk mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively
4.	Figure 4	Analysis using Random Direction	Figure 4, represent the results generated on the basis of mobility model namely Random Walk mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively
5.	Figure 5	Analysis using Random Waypoint	Figure 5, represent the results generated on the basis of mobility model namely Random Waypoint mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively

6.	Figure 6	Analysis using Cluster Movement	Figure 6, represent the results generated on the basis of mobility model namely Cluster Movement mobility model against the routing protocol Epidemic, Prophet, Spray and Wait, HBPR and GAER respectively for the network parameters packet delivery, packet overhead and average latency value respectively

S.No.	Abbreviations	Full form
1.	DTN	Delay Tolerant Network
2.	WAN	Wireless Area Network
3.	OppNet	Opportunistic Networks
4.	MANET	Mobile Adhoc Network
5.	ONE	Opportunistic Network Environment
6.	kRoP	K means Routing Protocol

Author's profile

RAHUL JOHARI

Rahul Johari is an Assistant Professor in at the University School of Information and Communication Technology (USICT) in Guru Gobind Singh Indraprastha University (GGSIPU), Dwarka, Delhi. His research interests include: - SHIP network, Cryptography, Security, IoT, Cloud Computing in which he has numerous SCI/SCIE publications. He is Head and founder of SWINGER (Security, Wireless Network Group of Engineering and Research).

TANVI GAUTAM

Tanvi Gautam is as Assistant Professor in IPEC, Ghaziabad and a research scholar of GGS Indraprastha University (GGSIPU), Delhi. Her research interest is in Cryptography and Opportunistic Networks in which she has 8 publications which includes book chapters and publications in ESCI journal and Scopus indexing.