EFFICIENT COOPERATIVE CACHING IN DISRUPTION TOLERANT NETWORKS

A DISSERTATION

Submitted in partial fulfillment of the requirements for the award of degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

By

SANDEEP KUMAR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

ROORKEE – 247 667 (INDIA)

MAY – 2016

DECLARATION

I declare that the work presented in this dissertation with title "**Efficient Cooperative Caching in Disruption Tolerant Networks**" towards the fulfillment of the requirement for the award of the degree of Master of Technology in **Computer Science & Engineering** submitted in the Dept. of **Computer Science & Engineering, Indian Institute of Technology, Roorkee**. India is authentic record of my own work carried out during the period from July 2015 to May 2016 under the supervision of **Dr. A.K. Sarje, Emeritus Professor, Dept. of CSE, IIT ROORKEE**. The content of this dissertation has not been submitted by me for the award of any other degree of this or any other institute.

DATE: ……………….. SIGNED: …………………….

PLACE: ………………. (SANDEEP KUMAR)

CERTIFICATE

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief.

DATE: ……………….. SIGNED: ……………………

Place: Roorkee

(Dr. A.K. Sarje)
Emeritus Professor
DEPT. OF CSE IIT ROORKEE

.

ABSTRACT

Opportunistic networks or Disruption Tolerant Networks (DTNs) are mobile networks in which a complete end-to-end path rarely exists at a time. Due to the high mobility of nodes, and limited radio transmission range, two nodes may not always be able to communicate with each other. Therefore, these networks are also called *Intermittently Connected Networks (ICNs)*. Thus, communication is plausible only at the encounter opportunities between nodes. This necessitates the use of *store-carry-forward* communication paradigm for routing messages from source to destination, in which intermediate mobile nodes store data to be transmitted till they find an appropriate relay node in the path towards the destination.

In Disruption Tolerant Network (DTN), determining the exact location of data and amount of delay to query the data by a requester is a major concern. It is costly for a node to maintain information of opportunistic paths to every other node in a DTN. Identifying appropriate caching locations is a difficult task. In this thesis, we propose a Greedy technique for selecting the nodes for caching data based on the past performance of the respective nodes so that data queries can be satisfied with less delay. Our basic idea is to cache data at a set of Proper Nodes (PNs), which can be easily accessed by other nodes in the network. We present a novel algorithm to select the PNs. This algorithm ensures that the PNs are not clustered to damage the data access performance of the whole network A Least Recently Used cache replacement method is use for replacing the data from the buffer of caching nodes, once the buffer is filled the caching nodes cache the data generated by source node and forwards the data to other caching nodes. A requester node broadcasts the data query to the neighbors and a caching node replies the query. Simulation results show the efficacy of the proposed approach in terms of query successful ratio.

ACKNOWLEDGEMENTS

I would never have been able to complete my dissertation without the guidance of my supervisor, help from friends, and support from my family and loved ones.

I would like to express my deepest gratitude to my supervisor, **Dr.A.K Sarje**, for his excellent guidance, meaningful insights, and moral support. He has been supportive since the day I began working on this dissertation and gave me the freedom I needed to explore this area of research on my own, while pointing me in the right direction in the times of need. His comprehensive knowledge in the area of Delay Tolerant Networks and hard working nature has been a constant source of inspiration.

I am also grateful to the **Dept. of Computer Science, IIT-Roorkee** for providing valuable resources to aid my research.

I would like to thank **Deepak S.** who supported me, were always willing to help and give me their best suggestions.

Finally, hearty thanks to **my parents and siblings**, who encouraged me in good times, and motivated me in the bad times, without which this dissertation would not have been possible.

Dedication

To my parents, for giving me the best education they could

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Disruption tolerant networks (DTNs), is a wireless mobile networks, in which nodes are mobile nodes in the DTNs contact to each other opportunistically. Due to the lack of connectivity, minute node density, random node mobility, all the nodes in the DTNs are intermittently connected. There is a difficulty of maintaining the persistent end-to-end paths between the nodes. Hence, the necessary requirement of the DTNs to use "carry-and-forward" technique for data spread. The basic problem in the DTN is how to decide the best relay nodes to store the data. The most vital application of the DTNs, "information *on-demand system*." This system used to provide the instant information to a node, which asks for precise information, like the "video on-demand system". There are so many forwarding scheme has been proposed in DTN, but there is partial work has done on proficient access of data. "*Cooperative caching*" is the ordinary method used to advance the access performance of data. Here we have learn about the "*cooperative caching*" and I proposed a new cache node selection algorithm, which used to select the caching node in the DTN and enables that elected nodes are not clustered and contact repeatedly to each other .

1.1 Overview

In the *Disruption Tolerant Networks (DTN)*, if nodes desire to talk in that case, there is opportunity that they are never meet each other. The short communication services based on the DTNs has turn into very well liked these days since, it enables a new suitable way to substitute information. Delay tolerant network, mark by their mobile nodes with their contact sequence in order to exchange the packets. Although many forwarding scheme have proposed in DTNs, However accessibility of efficient information is also very important.

1.1.1 Definition of DTN

Disruption Tolerant Network (DTN) also known, as Opportunistic Networks are type of networks in which the link between the nodes on the network is lively, which means the end-to-end path between the nodes is not persistent. Disruption Tolerant Networks have used in *Pocket switching networks* (PSNs), *Vehicular Networks Sensor Network* , *Ad hoc Networks* (MANET), etc .The data in the networks is distributed. Because of the accidental network topology, the connectivity between the nodes, the bandwidth and connection time are unbalanced. Due to uncertainty, it is necessary that in the DTNs that use the *store-carry-forward,* technique *to* ensure the accessing of the efficient data. Figure 1.1, Because be short of end-to-end connectivity, The Routing in the delay tolerant network is challenging job. The links in the network are depends on the time, there is no persistent topology of the networks, and to discover the subsequently node in the network for relay, nodes will store the messages until it find the paths. To decrease the delay in the delivery of the messages is very difficult.

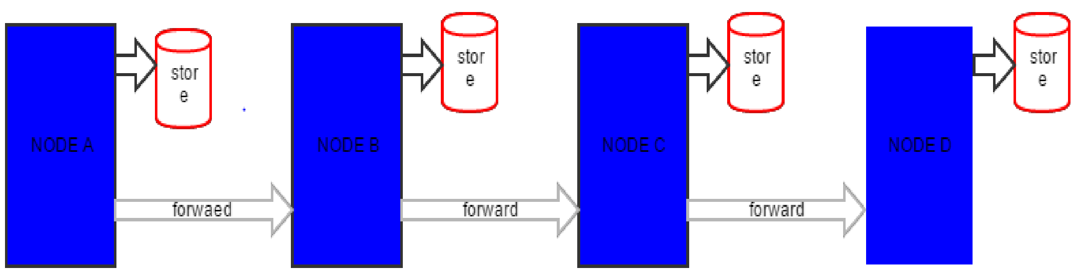

Figure 1.1 DTN Forwarding Environments

1.1.2 Application of the DTN

DTN's are use by the users in daily live such as humans,vehicales, agents etc. fig 1.2. In the section we are going to discuss about varies type of application of the delay tolerant networks(DTN).

DTN has the many application in *Social-based Mobile Network* in which social networking service are provided without internet 'Bytewalla'[1] uses the smart phones for exchange of packets when the internet's existing network can't be used for the emargency purposes.,like when the internet got shut down. It uses stor and forward technique and a sentinel surveillance helth-care application(SSA), it implements the bundal layer protocol in the IP architecture, which makes the communication easier .D-Blook[2] is a mobile application of social-based mobile network which allows users to creat, modify and share profile . it ia an implementation of DTN on .NET platform which can be used on various platforms,like windows,etc. Different small application about user's interest and their contacts.it also allow the users to searches the profile of different users based on therir interests. It also allows users to send massages to other users andsubscribe to the profile of other users. GeoLife [4] is a location-based social network, which uses GPS data driven networking. that forms the network of the people according to their location histories to provide various services such as Sharing Life Expriences , Generic Travel Recommendations , and Personalized Friend and Location .

In "*wildlife tracking"* DTNs is used for tracking the activity of the animals in the wild life , here we used the peer-to-peer wirless networking system. "Saami Network Connecivity(SNC)project[5]" uses DTN to establish Internet communication for the saami populition of reindeer Herders,who live in remote areas in Swedish Lapland ,so that they can relocate their base location according to the behaviour of the Reindeer for the environment protection. The Zebranet project [6] has installed a global positioning system(GPS) in a zebra collar to study the habits of the zebra activities and their migration . They have discussed efficient techniques for the power supply to the sensor networks and how to manage the energy consumption of the nodes. SWIM[7] is an underwater delay tolerant communication network , which is used to monitor the whales in the ocean. It tries to improve the capicity of the mobile network, with the trade off the delay . MULEs[8]. Is also used to collect the wildlife data . it gives a three-tier structural design for the collection of sensor data in less dense networks. MULEs are the mobile entities which are present in the environment, they, when in the range of sensors ,can take the data from them and deliver it to the desired point . this helps in the power saving of the sensors, which reduce the cost of the communication .

DTN also use in the field of communication between small communities . Daknet[9] is used to provide asynchronous digital connectivity in village communication , which reduces the cost of the communication. Vehoicular Wireless Burst Switching(VWBS)[10] is a DTN-based architecture which is used to give connectivity with low-cost for connectionwhich does not have any internet connection and doesnot have infrastructure of network. In this given architecture, automobiles act as mobile routers and they carry the packets between e-kiosks in the remote regions. E-kiosks are stations which receive the packets; they can be an educational , entertainment or halth unit.

DTN can also be used for providing communication services between moving vehicles.'Thedu'[11] gives a technique for the interaction between the automobiles for the searching of web page and sharing of them with each other in the areas where there is low internet connectivity, which makes web search possible from a vehicle with the help of proxy . The proxy store the search engine results and also the pre-fetched webpages and the mobile nodes are able to access the web pages from the proxy . CarTel[12], a VANET , is a distributed sensor communication system is designed to collect, transmit, and visualize data from sensors, located on mobile units of vehicles. A CarTel node is made up of sensors. It provides a query-processing interface ,which manages different types of data from sensors . The mobile node and the portal of delay-tolerant continous uery processor is made up of a delay tolerant network stack.

Moreover, in future DTN can also be used for networking between the setellites[13][14] . The satellite communication has a long delay, less end-to-end connectivity and also suffer from packet loss. Therefor, the TCP/IP protocol can not be used for the satellite communication. DTN architecture , which specifies the bundal layer protocol, can be used for satellite communication , which can address the above issues. Application of the DTN in mobile health systems and mobile , disconnected networks can be found in[16]respectively .

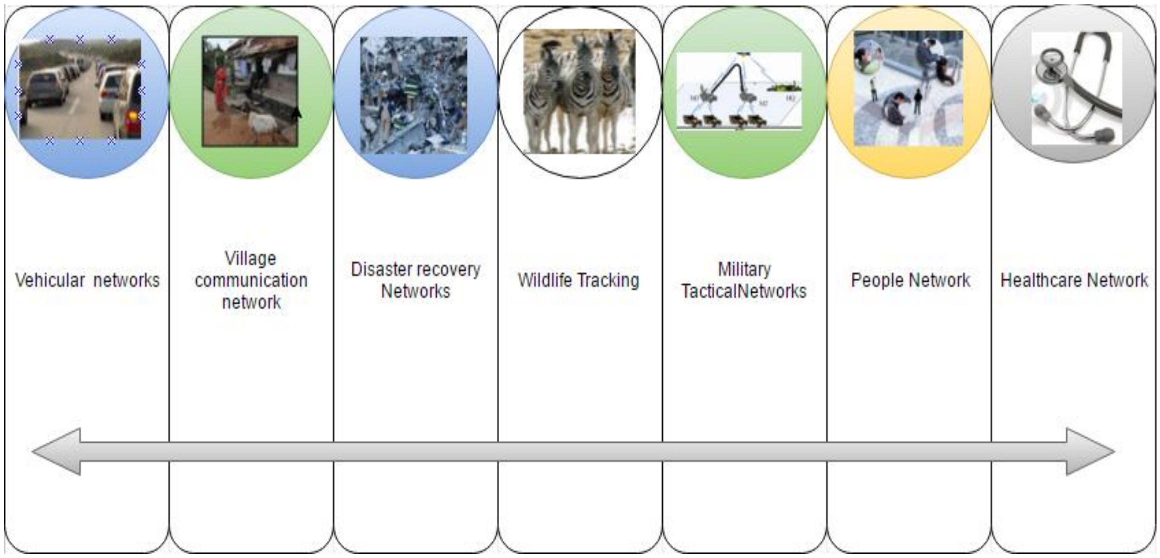

Figure 1.2 Core Application of the DTN

1.1.3 Challenges in DTN

There are different kinds of challenges face by the Disruption tolerant networks:

- **Mobility of the Node**: The network consist of mobile nodes ,as the devices carried by the people are typically mobile devices.
- **Frequent Network partitions**:Due to the frequent node mobility of the nodes it can be likely that there is no path available among the nodes**.**
- **Topology of the networks** : The network topology is not well-known earlier.
- **Energy Level Constraint**: The devices in the DTNs may have low energy .
- **Space of the Buffer**:Nodes in the network comprise small memory storage for holding the communication information .
- **Latency is Very High**: The latency between the nodes in the DTNs is very haigh , because there may be chance that two node may never meet to each other . this is also a big issue.
- **Larger Queuing Delay**: It is the consequence of regular disconnections in the network.

1.2 Routing Strategy in DTN

Routing in the can be done with the help of social characteristics of the nodes, like friendship, community,etc; or by forcing the nodes to participate in the routing process by giving credit to them;or it can be done just by mere exchanging the packets when encountered with other nodes . Based on these different routing behaviors the routing strategy in DTN can be classified into.

1.2.1 Routing with Social Metrics

The social based routing scheme based on the societal characeteristics knowledge of the network to take routing decision. These characteristics ard stands for long period and because of the node mobility it is not much livel. There are different kinds of social characteristics as , node closeness, Because probable all the nodes in their group meets frequently with the associated group nodes . Degree of the node is also an social metrics for routing , because all the nodes in the same group having the similar interest , and shaving the Chance for meeting in the nearest opportunity. The node with high deegree is a good option for next relay.

1.2.2 Opportunistic Routing

In the opportunistic routing , the packets are forward by the nodes one time two nodes get together. The packets are forwardwd with expectancy of high delivery . It is a flooding based system , in this system multiple copies of the same messages are flooded by the source in the network. The selection in this routing done on the bases of that these nodes will have the further contact to each other. Some of the forwading algorithm are discuss below which we have used in this project.

- **Epidemic Routing**: The goal of the Epidemic Routing[15] is to deliver messages with high delivery probability to the destination node. As it can be seen from figure1.3, that the source of the messagesse floods the messages to C1 and C2 ,which are in its range; C2 then stores the message till it meets another node C3.then C3 stores the message and finally deliver it to the destination D. So overall goal of epidemic routing is escalate the delivery rate and minimize the delivery latency.

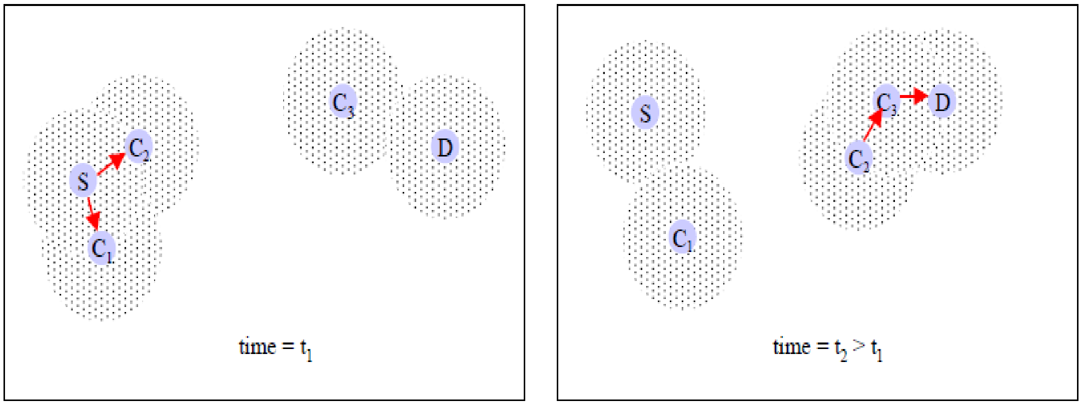

Figure 1.3 Epidemic Routing

All nodes are having unique ID. Nodes in the networks having the information about messages and network in its summary vector SV. When two or more nodes come into the range of each other, the node with lowest ID initiate a session that is called the anti-entropy session with the other node. After this session all the contacted node are having the same summary vector. The some messages during this session may find their destination.

- **Spray and Wait(SnW)**: Spray and wait algorithm comprises two phases (spray and wait) in the spray phase a node sends initial L coppies in the network to the nodes with which it hase direct contact with .In the wait phase when the destination is not discovered in the spray phase , the sender node hase to wait till the initially sprayed copies meet the destination. In the case of Binary *spray and Wait* , if a node has multiple copies of a packet and if it has a connection with node B , then the former node gives half of the copies to the later node and keeps halfe of it . And when a node is left with one instance of the packet ,it can do only transmission to destination node.

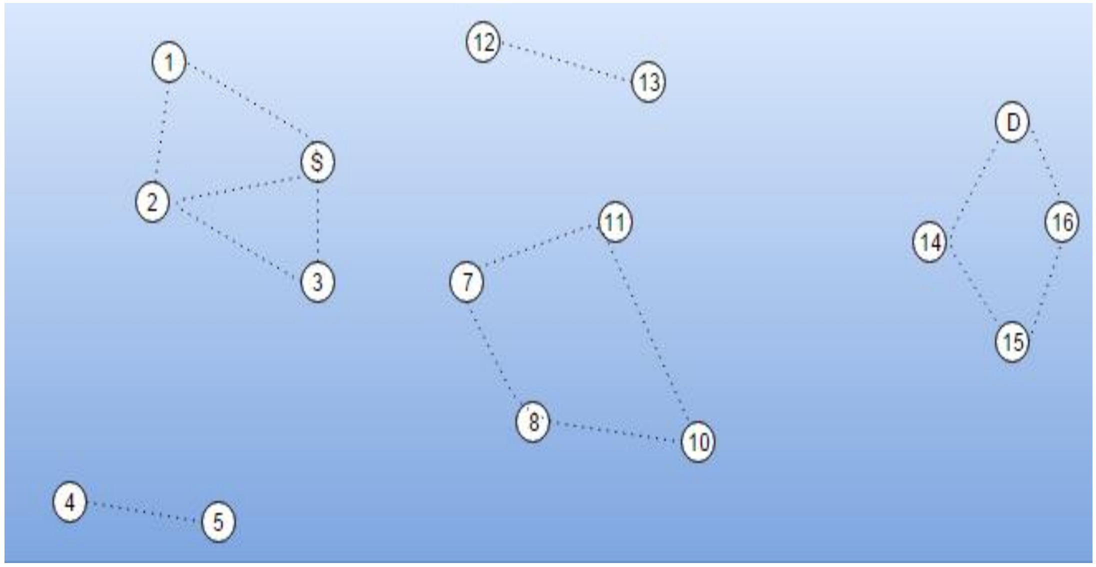

Figure 1.4 Spray and Wait Routing

We can see an example in figure1.4, in which node S sprays three copies to node 1,2 andrespectively. Node 7 sprays 2 copies to node 11 and 8 and is left with half of the copies. Node 11 and 8 send their copies to node 10and so on. Node 14 and 16 is the node who were finally left with only one copy and they send it directly to the destination D.

The advantages of this algorithm are it has the less number of transmissions when compared with other flooding-based schemes ,when the load is high in the network,it has the less delivery delay. Its also scalable, as the number of nodes in the network increses the number of transmissions per node decreases.its very simple algorithm and requires no knowledge of the network.

- **MaxProp**: MaxProp prioritizes the packets which it has to send the peers . In MaxProp it is assumed that each node has unlimited buffer . A node can send the packets to any number of nodes until the TTL of the packet expires. It has the ranked list of the neighboring nodes , cached messages which is based on the cost which is assigned to every destination. The cost is delived from the delivery probability. The packets which are new are assigned a higher delivery probability.

  In the MaxProp each node iϵS has a probability of meeting a node jϵS . this probability can be represented as f_j^i , which is the probability of encounter of the node i with node j. Initially the value of f_j^i is initialized as 1/(S-1) where S is the total no of node in the network. When the node I meet node j, the value of the f_j^i is incremented by 1 and then all the value of f are normalized.

- **Prophet** : It is a guess based routing use the past of meet for guess. it is used to reduce the Epidemic Routing paths. The objective of this routing protocol is to transferring bundles only on routes that have high likelihood to reached at the target.

To assess the efficiency of the routing algorithm , different types of performance metrices can be used some of them are , Delivery ratio, Delivery Latancy, and Delivery Cost etc.

1.3 Data Access Technique in DTN

Caching is is the most popular stratergies which is used in different application of networking for data access , there are two types of caching used in the field of data access , In the network.

1.Incidental caching and

2. Intentional caching .

In the intentional caching the data is store at the specific set of the node ,but in the caes of incidental caching the intermediat node cache the pass-by data. The proper node(specific node) or in the intentional caching is called the central node which is choosed based on the some additional information ,like path weight , number of neighpors ,contact pattern ,contact rate, contact duration ,number of past contact etc.

In the case of Incidental caching the caching node are choose based on the history, so that it would fail in the case of opportunistic contacts and intermittent connectivity (DTN), Hence In the case of DTN intentional caching are used.

Some other caching scheme comprised, cache path, cache data, distributed caching, and cooperative caching. Cache path is storing the path towards data source, while cache data is the simplest scheme in which all nodes caching the pass-by data rather than path. In the distributed caching and cooperative caching, data being stored at some places are coordinated while generating responses. i.e Probabilistic responses are given in order to overcome transmission of the duplicate information , these scheme work well in case of normal Ad-Hoc networks but they fail in the DTN because of the, High mobility and assumption of request- response follows the same path in AD-Hoc network fail in DTN. Depict in figure1.5.

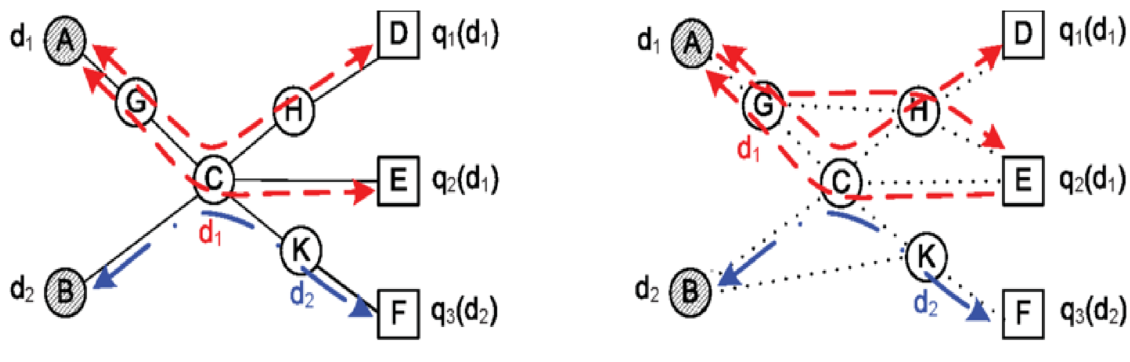

Figure 1.5 Caching strategies in different network environment [19]

In the Figure 1.5(a), all the request are forwarded to the sources of the data A and B , and also data $d1$ and $d2$ are forwarded to the requesters. Due to the partial cache buffer, C is only cache popular data $d1$ based on the query record, and likewise node K store the data d2. The pass-by data incidentally cache by the any node in the system generally.

However, In the DTNs due to the frequent partitioning, the caching of the data incidentally is not effective in DTNs. In the DTNs data, forwarding path is opportunistic, So that request and response follows the different routes. Due to these intermediate nodes are not able to accumulate the knowledge about the queries history and making decision of caching. For Example, in Figure 1.5(b), The source A of the data having query q2, and link between C and G may drop, and hence node C does not able to cache data d1 replied to E. Source forward the data to node H, and it does not store the pass-by data d1 either since it did not have the past request and consider d1 as a less popular query $q2$, the link between C and G may be down , So that node C does not able to cache the data d1 replied to E. The source node now having the link with node H and its hand over the packet to node H, while its does not store the data d1, because it consider the data d1 either less popular or it does not have the query q2.

1.4 Research Gap

In the DTN, all the existing cooperative caching scheme cache the data at the Intentionaliy caching node that is called the *network central location (NCL)* but they did not consider the situation of contact among the selected NCLs, and condition of the query of different kinds of data items . In the field of data, forwarding many researches has done in DTNs field. However, in the field of efficient data access limited work has done.

To best of my knowledge, there is no analysis has been done till now, on the contact situation of the selected proper nodes (PNs) an and queries condition for different kinds of data items

1.5 Research Significance

- The propose research gives the efficient data access heuristic in the Delay tolerant networks.
- The outcome of our research will be a simple delay mitigation technique, which can be use to increase the performance of data access, optimize the tread off accessibility of the data, and query successful ratio.

1.6 Problem Formation

In the DTNs, the data access delay between the nodes some time is very high with compare to the traditional internet, due to the intermittent connectivity between the nodes of DTN.

Given : A homogeneous delay tolerant network , in which a packet P can pass from source S to destination D , Each node have the limited energy resource and buffer space and its having the infinite computational power , the contact time is consider enough to pass all require data from one node to another node. Each node having the unique identification ID

- **Constraints**: It is considered that the data access delay between the pair-wise nodes is very high due to this successful ratio is very low.

- **Objective**: Our basic idea to reduce the access delay between the pair-wise and improve the queries successful ratio.

1.7 Dissertation Overview

The dissertation comprises of five chapters:

- Chapter 1 introduces Delay Tolerant Networks, their applications and challenges. We briefly formulate the problem and summarize the contributions made in this dissertation.
- Chapter 2 reviews the work related mainly to caching scheme in the delay tolerant networks.
- Chapter 3 comprises of the novel approach of selection of the caching nodes in the delay tolerant networks.
- Chapter 4 including the simulation result after the caching factor applied to the opportunistic algorithm. In addition, discussed about the results after applying the greedy method to resolving the proper node selection condition.
- Chapter 5 summarizes the dissertation and provides some open problems in this area for future extensions of this work.

CHAPTER 2

LITERATURE REVIEW

In chapter, we have discussed about the related work that has done so far in the area of data access in Delay Tolerant Network (DTN). Data access in DTNs can provide in different ways. Data based on the interest profiles is distribute to intended user, the most commonly used system for data distribution is Publish/subscribe systems, and they generally take advantage of structures of the social community to find out the relay.

Caching is a different technique to provide access of data. We have studied *Cooperative Caching in Wireless ad-hoc networks*, intermediate node in the cooperative caching in mobile ad-hoc network store the pass-by data on the past request and request in future responded with less delay.

2.1 Motivation

The diverse cooperative caching schemes deployed in the DTNs. Reisha et al [21], anticipated adaptive caching scheme based on Learning Automata (LA). The DTN environment knows the scheme for LA's adaptability, while it does not deal with load balancing issues.

One more cooperative caching scheme anticipated by Xuejun et al [22], is called as duration aware caching by considering the common parameters of the network. This method claims the ease of data access when the data is fragmented based on a social parameter called 'Contact Duration (CD)' among nodes.

In [25], Yin and Cao, anticipated a caching protocol in mobile ad-hoc networks, in data and the path of the data store at a particular relay node, this is helpful in reducing the response time of the future request. However, due to the intermittent connectivity between the DTNs nodes it is not useful in the DTNs

Wei Gao and Guo hong Cao[29] anticipated an efficient method which ensures appropriate network central location selection based on a probabilistic selection metric, and furthermore coordinates various caching nodes to optimize the tradeoff between data accessing and overhead of the caching .

To best of our knowledge, there is no analysis has been done till now, on the contact situation of the selected proper nodes (PNs) and queries condition for different kinds of data items. Now in this report we have proposed the novel algorithms for selection of the caching nodes in the DTNs environment, which consider the situation of the contact between the DTNs nodes.

2.2 Summary

In this chapter we have discussed about the theoretical analysis that have been done about caching in *Delay tolerant networks* .we classified all the work that has been done about the caching in the DTNs. So, from this chapter, we conclude that limited works has been done about the caching in the DTN. In the next chapter, we will describe our proposed work, which comprised of how to select the proper node in the challenging networks (DTN) and how to include it in the network, a novel scheme for selecting the proper node or central node.

CHAPTER 3

PROPOSED APPROACH

The first section of this chapter discuss the networ model and highlite the inspiration of the caching in the DTNs. The second segment of this chapter discuss about the proposed scheme we have used in proper node(PNs) selection in the DTNs.

3.1 Network Model and PNs Selection Metric

The network is considered as a graph G = (V, E) containing V set of vertices, called nodes and the set edges E is called the stochastic contacts. We suppose that the links are symmetric, .and due to the node mobility and limited range of the communication the contacts are opportunistic and network environment is unstable. Each node is assumed to contain a certain fixed amount of energy. The nodes lose energy at a constant rate during message exchanges. Each message is assumed to have a fixed TTL (time to live) value after which it expires. All the messages are of equal priority initially.

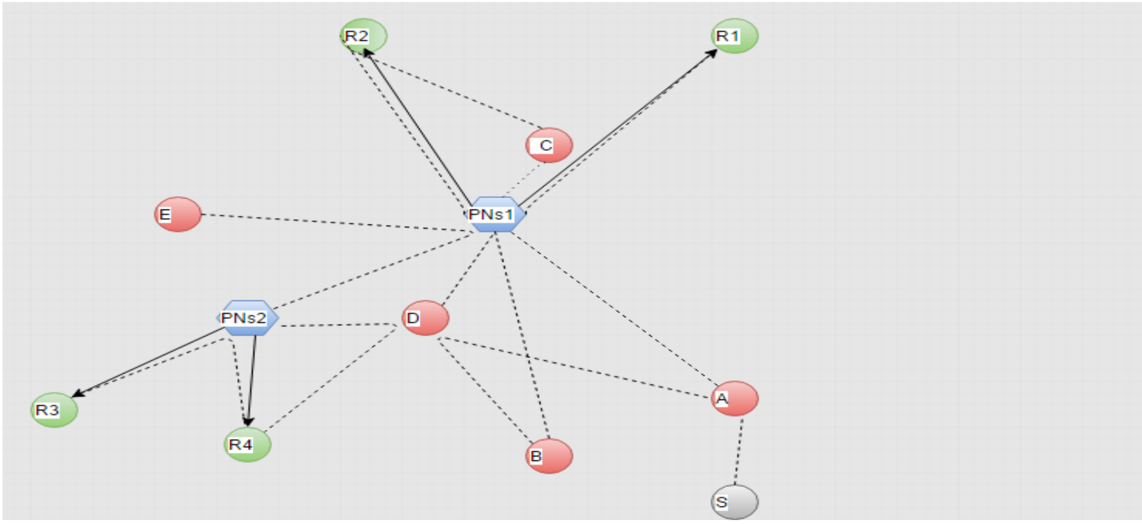

Figure 2.1 Disruptions Tolerant Network Model

In the above model figure 2.1, the nodes are represented by the circle and hexagon, Opportunistic connection between the node is indicated by the thin dash line. Where nodes R_1, R_2, R_3 and R_4 are the data requester nodes, the S generate the data. The source of data S, directly forward the data to requesters, if there is no cache in the system and total no of hope is 13. if PNs1 and *PNs*2 selected as relays nodes ,then source of the data forward the copies of the data to proper node PNs1 , and PNs2 copies data from PNs1. Then the selected proper nodes respond the all requesters node. In the above diagram now, total number of hope count is seven. The above summary shows that if data is cache at some proper nodes, then its take the fewer number of hops. Which can improve the successful ratio or data access performance.

There is some challenges face by the proper node selection method in the DTNs. First, the selected proper nodes should not be clustered, because the clustered PNs reduced the functionality of caching. Second, the communication between the selected proper nodes should be easy; Because of this, they can response the queries frequently. There are different kinds of matrices used in PNs selection.

- **Contact Rate**: The inter contact time between any node N_k and N_{k+1} in the path follows the exponential distribution with contact rate λ , The contacts rate between pair-wise node can be calculated based on the connectivity record between the nodes. The contact rate between nodes $N_i and. N_j$ is $\lambda_{i,j}$ as follows.

$$\lambda_{i,j} = K/(T_{\Delta+k} - T_{\Delta}) \qquad (1)$$

where k denote the number of latest contacts among the node $N_i and. N_j$, and T_{Δ} denote the starting time of the history and $T_{\Delta+k}$ denote the ending moment of the recording. Using these parameter we can find out the contact rate matrix $A_0(G)$ between the pairwise nodes.

$$A_0(G) = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & . & . & . & \lambda_{1,n} \\ . & . & . & . & . & . \\ . & . & . & .. & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \lambda_{n.1} & \lambda_{n,2} & . & . & . & \lambda_{n,n} \end{pmatrix}$$

The contact rate between pair of node is symmetric , so the matrix should be symmetric matrix , that means the sum of the columns and rows should be equal.

Because there is no fixed infrastructure and there is no global knowledge about the networks, every node maintains the information about the pair-wise contact rate and shortest opportunistic path to other node. The probability density function (PDF) of the inter-contact time X_l between node" N_{l-1} *and* N_l is.

$$f_{x_l}(x) = \lambda_l \ell^{-\lambda_l x} \tag{2}$$

Here we assumed, the symmetric contacts rate between every pair of node in the network, in order that $\lambda_{i,j} = \lambda_{j,i}$ describe the contact-rate between N_i and N_j. If, there is no direct contact between the nodes N_i and node N_j. Then the contact rate between them $\lambda_{i,j} = 0$ for all the nodes in the set V, $\lambda_{i,i} = 0$. We have to finding the path between the pair-wise node with high contact rate and with less hops its Our primary goal.

If there is no direct contact between the node pair then we have to find the multi-hope opportunistic contact path between the pair-wise nodes in the contact network. A r-hop opportunistic path P_{AB} between nodes A and B consists of a node set V_P = (A,V_1, V_2 . . .Vr-$_1$, B) $\subset V$ and an edge set E_P = (e_1, e_2, e_3,……e_r)$\subset E$ weights of the edges are (λ_1, λ_2, λ_3----- λ_r). The weight of the path $P_{AB}(T)$ indicate the probability of that data transmission opportunistically from A to B along P_{AB} in time T.

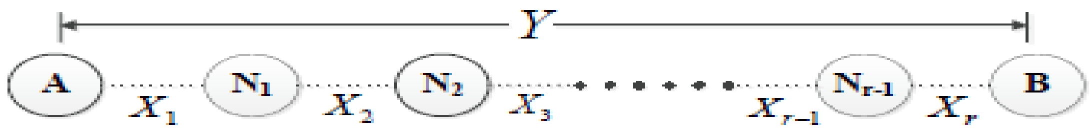

Figure 2.2 Multi hop Opportunistic Path

Where X_1 denote the inter-contact time between A and N_1 , and X_2 is the inter contact time between N_1 and N_2 similarly the inter-contact time linking the node N_{r-1} node B is X_r. the total inter- contact time between A and B is $Y = \sum_{l=1}^{r} X_l$.

We have to consider the contact between the different pair–wise node is independent. Hence the contact time between the node , X_l is exponential random variable with respective contact rate λ_l (l=1,........r) , and $\lambda_i \neq \lambda_j$ for $i \neq j$. the total inter contact time between the A and B follows the hyper exponential distribution[20] and the probability density function of it .

$$f_y(x) = f_1(x) \otimes f_2(x) \otimes .... \otimes f_n(x) \qquad (3)$$

$$= \sum_{l=1}^{n} M_n^l \lambda_l \ell^{-\lambda_l x}$$

Where $M_n^l is$

$$M_n^l = \Pi_{1 \le q \le n, q \ne l} \lambda_q / (\lambda_q - \lambda_l)$$

The weight of the opportunistic path from A to B is defined as follows.

$$P_{A,B}(T) = \int_0^T f_y(x) dx$$

$$= \sum_{l=1}^{n} M_n^l \ell^{-\lambda x} \qquad (4)$$

The data spread delay should be fewer than T connecting the node A and node B. for the selection of the Its PNs node we need to find the weighted contact path of every pair-wise node and based on this analysis we have to calculate the matrix $C_1(G)$ of weighted opportunistic path between pair of node . The weight of the opportunistic path between pair of node we have calculated using Eq.(4). Because of symmetry, the weight of the opportunistic paths is symmetric

$$C_1(G)=\begin{pmatrix} P_{1,1} & P_{1,2} & . & . & P_{1,n} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ P_{n,1} & P_{n,2} & . & . & .P_{n,n} \end{pmatrix}$$

between pairs of nodes, For example , The opportunistic path weight between node v_i and node v_j is $P_{i,j}$, So that $P_{i,j} = P_{j,i}$ for all the node $v_i \in V$. Matrix is symmetric so the sum of the rows is equal to the sum of the columns.

- **Interest level**: In this analysis, we have calculated the interest level of the node on the basis of the queries generated record, the calculation of this as follows.

$$q_{i,m} = h_m / t_{\Delta+h_m} - t_\Delta$$

In the above formula h_m is the number of request for the data item d_m created by the node v_i during the period $t_{\Delta+h_m} - t_\Delta$. The interest level of the node of the node v_i for the data item d_m denoted as $Q_{i,m}$.

$$Q_{i,m} = q_{i,m} / \sum_{i=1}^{n} q_{i,m}$$

Now using these matrices, we have to find the level of the interest of the each v_i for every data point $d_i \in d_m$. And now, we have to get a matrix R, and total number of rows in the matrix is number of nodes in the networks and total number of columns is equal to the number of data items and the entries in this matrix like this $q_{i,j}, 1 \le i, j \le n$ the total sum of the interest level of the all node for a particular data item, is 1 can say $\sum_{i=1}^{n} q_{i,K} = 1, \forall d_k$.

The multiplication of the matrix $C_1(G)$ with a row R(*,m) in the matrix R, getting a matrix M(G).

$$M(G) = \begin{matrix} v_1 \\ v_2 \\ | \\ | \\ v_n \end{matrix} \begin{pmatrix} p_{1,1}q_{1,m} & p_{1,2}q_{1,m} & - & - & p_{1,n}q_{1,m} \\ p_{2,1}q_{2,m} & p_{2,2}q_{2,m} & - & - & p_{2,n}q_{2,m} \\ - & & & & \\ - & & & & \\ p_{n,1}q_{n,m} & p_{n,2}q_{n,m} & - & - & p_{n,n}q_{n,m} \end{pmatrix} \quad (5)$$

In the above matrix $P_{i,j}$ denote the probability of the node v_i contacting with node v_j in time t, it obtaining from Eq.(4). In addition, $q_{j,i}$ query rate of a node v_j for data item d_m . Hence, $p_{j,i}q_{j,m}$ evaluates the condition that node v_j requires data item d_m and the node satisfies it v_i within the time T, if a node v_i have data item d_m. The average values of the cells in the particular column i of the M(G) indicate the metric for considering the Its PNs node , for node v_i for caching of the data point d_m, it is defined as below .

$$PNs_1(i) = 1/N\sum_{i=1}^{n} p_{j,i}q_{j,m} \quad (6)$$

PNs (i) indicate the average probability of a node v_i, That can transmit the data item d_m to an arbitrary requester with in T. A node, which has the largest value of PNs (i), select as a first proper node.

To optimize the metrics of the PNs selection bearing in mind the contact condition with the elected PNs, consecutively to assurance that the PNs should not be cluster. The clear declaration To make PNs selection is that , Its select the second PNs , Then we apply the greedy technique for proper node selection in DTNs.

To overcome the problem of the clustering of PNs, select the first PNs , after this , decreasing the contacts rate between first PNs and all other neighbors with a constant α ($0 \le \alpha \le 1$) and we got the matrix $A_1(G)$. That reduces the probability of the first PNs and the node, which has the highest probability, select as the next PNs. If all the PN's are cluster then it reduce the functionality of the caching now again apply the all step on the $A_1(G)$ and find the next PN Further more the value of the α depending upon the size of the data or the size of the buffer of the nodes. G in this experiment, we have considered α=.25 Greedy techniques for proper node selection discuss as follows.

3.2 Algorithm : Proper Node Selection

Input:

G(V,E): Disruption tolerant network .

$A_0(G)$: Matrix of contact rate between the pair-wise node

μ: The number of proper nodes.

α: A constant ($0 \leq \alpha \leq 1$)

K=1; index of the selected proper node;

Output:

Set of all selected proper nodes.

While K≤μ do

1.Find the path weight matrix $C_k(G)$.

2. Find the matrics $PNs_k(i)$.

3.$PNs_k = avgm_i ax\langle PNs_k(i) | v_i \in V\rangle$;

4. K=k+1;

5. Adjust $A_k(G)$ with the constant α.

6.End while;

3.3 Summary

In this chapter we have discuss about the proposed scheme, for increasing the delivery performance and successful ratio of the queries of the data. In the next chapter we have discuss about the tools and technology, we have used in this thesis.

CHAPTER4

SIMULATION RESULTS

This chapter including the simulation result after the caching factor applied to the opportunistic algorithm. In addition, discussed about the results after applying the greedy method to resolving the proper node selection condition.

4.1 Simulation Setting

The simulation setting of ONE simulator is given in the table 1, we have consider total 40 nodes in my experiment , the movement model is Random way point , we have executed the simulation for 12 hours deployed in an area of 4500*3500 m^2 . The networks warm up period is 6 hours. The simulation time is the total number of time spam, which have considered for node communication. The transmission speed of the node is the data rate at which packets are transmute from one node to another node. Time to live is the total time after which a packet get drop from the networks, the message creation interval is the time range in which a nodes creates packets for each other. Wait time is the time for which each node waits before moving to next location.

The First half of test is use to for gathering the information about the network like, rate of the contact, query rates, and select caching node. The duration of the first half of our experiment is 6 hours. Here we have considered that each proper node is corresponding to a different type of data. For example, six sets of proper node are corresponding to six types of data.

Table 1 Simulation Settings

PARAMETER	VALUE
Simulation Time	43200s= 12 hours
Total Number Of Nodes	40
Time To Live (TTL) of message	2 hours
Message Creation Interval	25-30 seconds
Transmission Speed of Nodes	250kBps
Interface	Bluetooth and high speed interface
Size of the data items	10KB
Buffer size of each node	1MB

We have to compare the performance of our scheme with the existing cooperative caching scheme [23]. We display the result of queries successful ratio “the query satisfies by the requested data within its life time”.

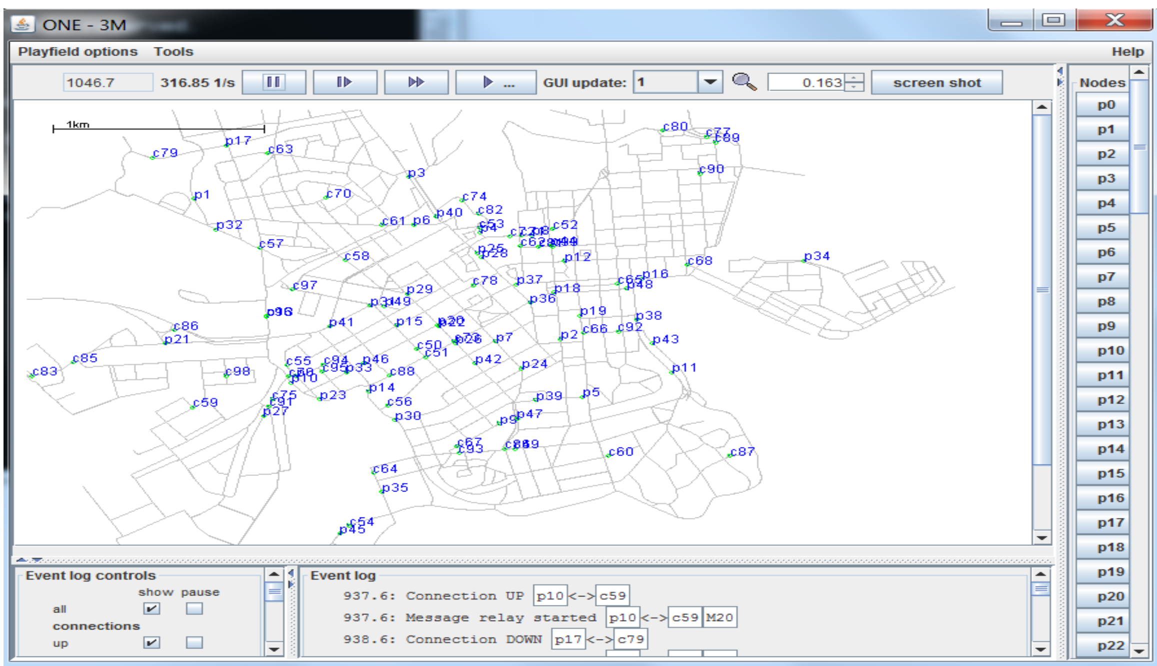

Figure 4.1 Screenshot of Simulation (ONE)

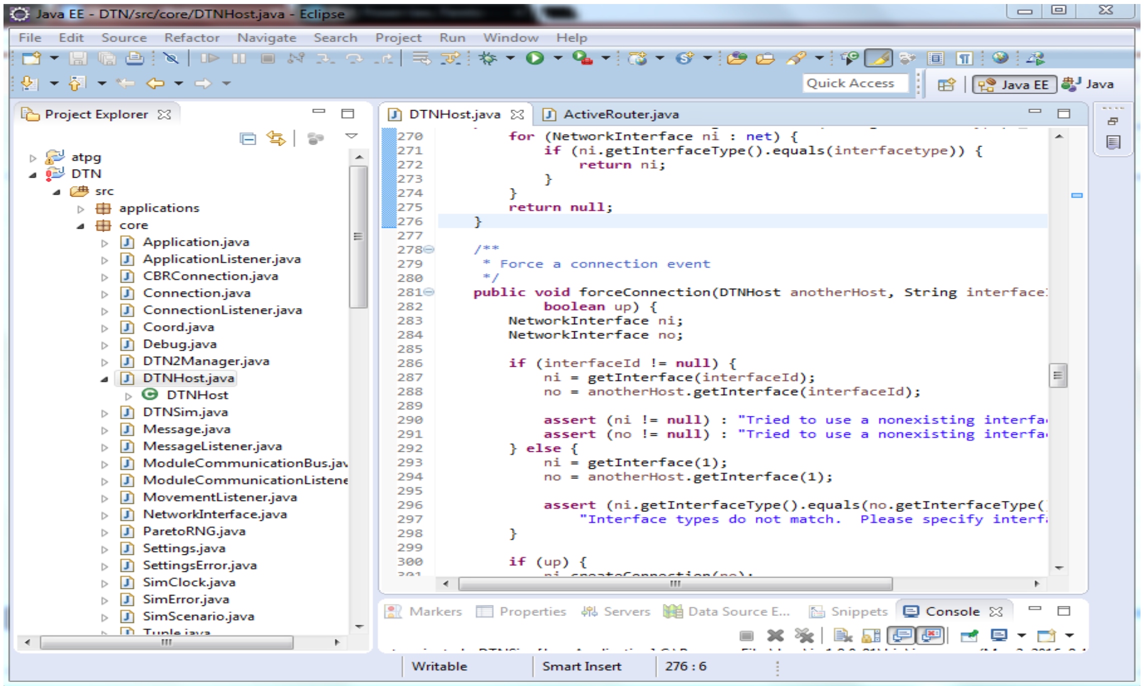

Figure 4.2 Screenshot of Eclipse

4.2 Performance Evaluation

In this section, we have discussed about the performance of our proposed, Proper Node Selection (PNs) scheme, and we have compared it, with presented caching scheme in [25]. The previous caching scheme , decide the appropriate network central location(NCL) to cache the data but they did not consider the query condition and contact condition in the selected , network central location.

In this experiment each having Bluetooth interface , each device created the queries for different kinds of data items, and when two device are meeting to each other the time of contacts was recorded . To analyzing the result of our experiment, we used the record of 6 hours for calculation of the “rates of contact and rates of query and selected cache node”. Second 6 hours we have calculate the data access performance. In this experiment, we have to select the nine kind of data and nine sets of the PNs corresponding to nine kinds of items of data. We have compared the query success ratio of the existing scheme to our proposed scheme.

The table 2 showing that, the successful ratio of our proposed scheme is incremented 12.32% with the existing scheme with respect to the lifetime of the data.

For example, the figure 4.3 showing that as increasing the lifetime of the data the successful ratio of the query will also increase. Each data packet having the lifetime T, the basic stagey of our project is that every request should be satisfied with in the lifetime of the data. The buffer of the node is having limited size, that why we have used the LRU as cache replacement techniques for this work. Each node in the networks caches the frequently used data, which increase the query successful ratio. When source broadcast the data, each node decide whether to cache the data or not, normal nodes forward the data to next neighbor and caching node cache the frequently used data.

Table 2 Successful ratio with respect to data life time (μ=7)

Life time of data(h)	No Cache	NCL	PNs
.6	5.09	19.76	20.23
.8	7.28	26.04	27.94
1	12.43	30.94	33.05
1.2	15.94	36.76	39.89
1.4	19.23	40.72	44.34
1.6	22.42	47.93	53.43
1.8	25.34	51.49	61.12
2	27.94	57.93	69.97

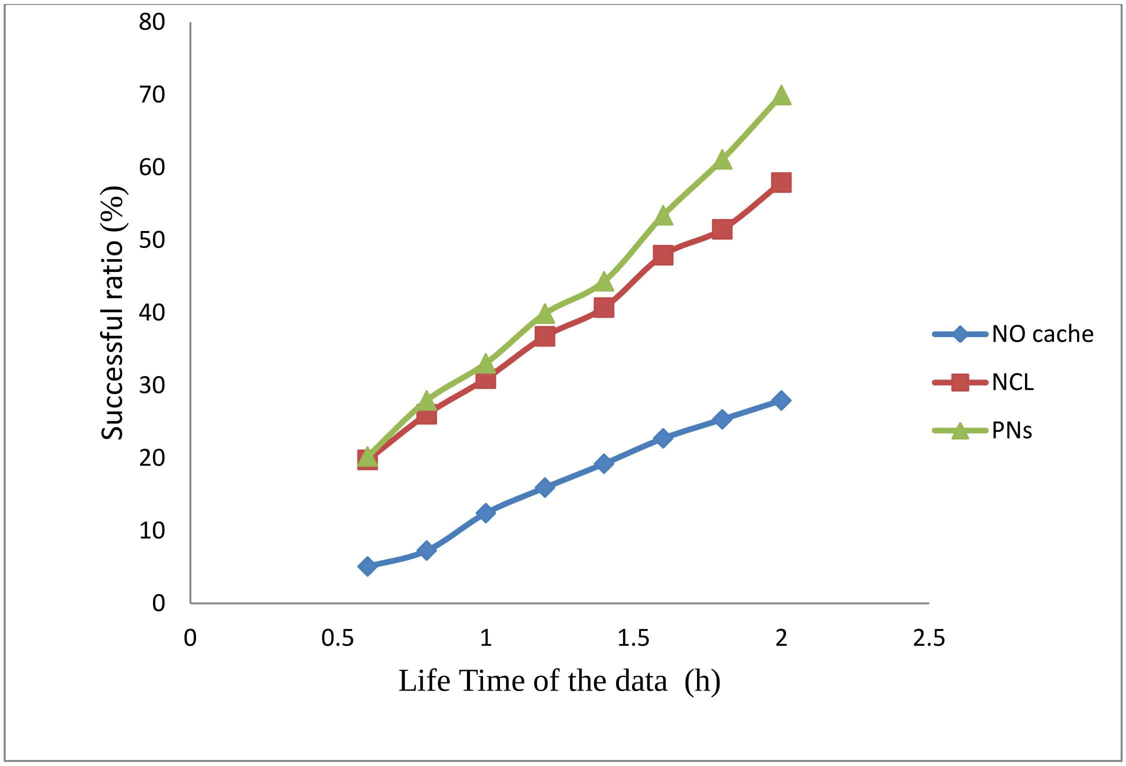

Fig 4.3. Successful Ratio With Different Data Life time of the data (μ=7)

Table 3 Successful Ratio with Respect to number of caching nodes

Number of caching nodes	NCL	PNs
1	18.44	20.33
3	26.38	27.56
5	29.54	33.94
7	34.43	37.69
9	40.34	43.12
11	42.33	43.94

- **Successful Ratio:** The ratio of queries being satisfied with the requested data.

The successful ratio of the query depending upon the lifetime of the data and the total number of the caching nodes in the networks, increasing the caching node the successful ratio also increase at the some extent.

The impact of the different number of the caching nodes is showing in the figure 4.4 when μ is reducing 3to 1 the successful ratio of the experiment decreases 35.56%. When μ reach at the threshold point then further increment in the μ will not increase the access performance of the data. It means that the larger value of the μ is not effective, so choose the optimum value of the μ. In the figure 4.4 after μ=9, the access performance of the data are approximate same.

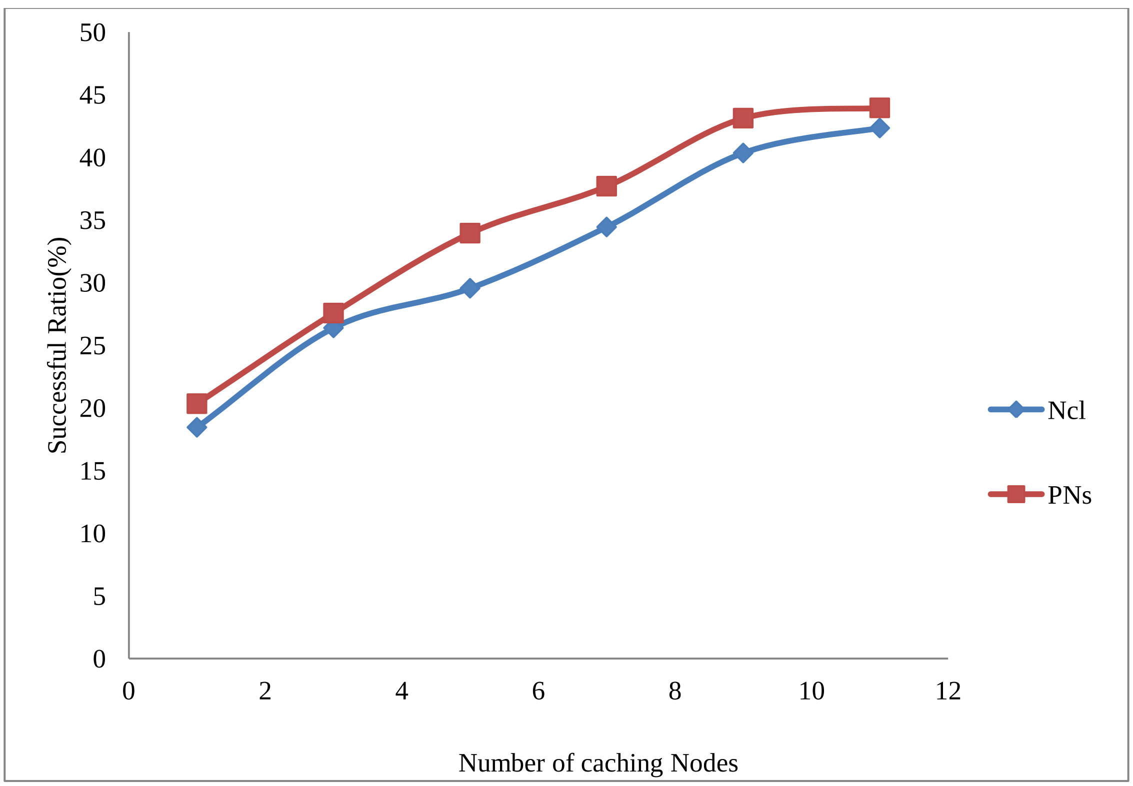

Figure 4.4 Successful Ratio with Different Number of Caching Nodes (Life-time of Data=1h)

4.3 Summary

In this chapter we have prove that our proposed scheme give the better performance then, the existing caching scheme. Here we have compare the successful ratio of the query, generated by the nodes in the networks for both the scheme. Our proper node selection (PNs) algorithm gives the better results. The result of PNs depending on the number of caching node in the networks and the lifetime of the data, at some extent our proposed scheme give the best result, but for larger value of the μ the impact of the proposed scheme would be minor.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this dissertation, we have analyzed the behavior of the caching scheme based on the Proper Node Selection algorithm , and we have present the greedy algorithm for proper node selection , by considering the contacts between the pair-wise nodes and query situation for different data item. We have evaluated our proposed scheme using the ONE simulator. At last, we compared the result of our proposed scheme with the existing caching scheme based on the successful ratio; the performance of our proposed scheme is 12.32% better than the existing caching. We have evaluated our proposed scheme using the ONE trace.

Research work can be extended in various areas. Some of them are:

1) Load Balancing

2) Cache Replacement

When the central nodes change, the existing caching locations become inappropriate, and hence, the successful ratio of data access may reduce. Load balancing is essential in the cooperative caching in DTNs. Load balancing of the caching in DTNs is a big issue. The cache replacement is also big issues in the DTNs, the traditional caching strategies, FIFO, LRU, are not efficient in the DTNs because of over simplistic consideration of data.

REFERENCES

[1] H. Ntareme, M.zennaro , and B.pehrson, “Delay Tolerant Network on smartphones: Applications for communication challenged areas.” In Proceeding of the 3rd Exterme Confererence on Communication , The Amazon Expedition,2011,PP. 14.

[2] R.J. Clark, E.Zasoski, j.Olson, M.Ammar, and E.Zegura, “D-book: a mobile social networking application for delay tolerant network ” In proceeding of the 3rd ACM workshop on Challenged networks, 2008,pp. 113-116.

[3] S.M. Allen, G. Colombo And R.M. Whitaker, “Uttering: Social Micro- blogging without the internet,”In Proceedings of the Second International workshop On Mobile Opportunistic Networking Networking ,2010,PP.58-68.

[4] Y. Zheng, Y. Chen, X. Xie, and W.-Y. Ma, “GeoLife2.0: a location-based social networking service,” in ... Conference on Mobile Data Management: Systems, Services and Middleware (MDM '09), pp. 357–358, May 2009.

[5] P. Zhang, C. M. Sadler, S. A. Lyon, and M. Martonosi, “Hardware design experiences in ZebraNet,” in Proc. 2nd Int. Conf. Embedded Networked Sensor Syst., New York, 2004, pp. 227–238.

[6] T. Small and Z.J. Haas, “The Shared Wireless Infostation Model - A New Ad Hoc networks the ACM MobiHoc 2003 conference, Annapolis, Maryland, 2003, pp 233-244.

[7] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks,” In ad-hoc Networks 2003 vol.1,no.2 pp.215-233..

[8] A. Pentland, R. Fletcher, and A. Hasson, “DakNet: Rethinking Connectivity in Developing Nations,” Computer, vol. 37, no. 1, pp.78-83.

[9] Farahmand F, Patel AN, Jue JP, Soares VG, Rodrigues JJ: Vehicular wireless burst switching network : enhancing rural connectivity,” In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM) WorkShops 2008: pp.1-7.

[10] A. Balasubramanian, Y. Zhou, W. B. Croft, B. N. Levine, and A. Venkataramani, “Web search from a bus,” In proceedings of the 2nd ACM workshop on challenged networks 2007, PP.59-66.

[11] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, “CarTel : A Distributed Mobile Sensor Computing System,” pp. 125-138.

[13] C. Caini et al., "Delay- and Disruption-Tolerant Networking (DTN): An Alternative solution for satellite networking applications,” In proceedings of the IEEE, 2011 vol. 99,no.11,pp. 1980-1997.

[14] A. Balasubramanian, B. Levine, and A. Venkataramani, “DTN Routing as a Resource Allocation Problem,” Proc. ACM SIGCOMM Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm., pp. 373-384, 2007.

[15] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web Caching and Zipf-Like Distributions: Evidence and Implications.” Proc. IEEE INFOCOM, vol. 1, 1999.

[16] E. Daly and M. Haahr, “Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs,” Proc. ACM MobiHoc, 2007.

[17] K. Fall, “A Delay-Tolerant Network Architecture for Challenged Internets,” Proc. ACM SIGCOMM Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm., pp. 27-34, 2003.

[18] W. Gao and G. Cao, “User-Centric Data Dissemination in Disruption Tolerant Networks,” Proc. IEEE INFOCOM, 2011.

[19] W. Gao, Q. Li, B. Zhao, and G. Cao, “Multicasting in Delay Tolerant Networks: A Social Network Perspective,” Proc. ACM MobiHoc, pp. 299-308, 2009.

[20] S. M. Ross. *Introduction to probability models.* Academic Press, 2006.

[21] Reisha Ali and Rashmi Rout, “An Adaptive caching Technique using Learning Automata in Disruption Tolerant Networks”, International Conference on Next Generation Mobile Apps, Services and Technologies,2014

[22] XuejunZhuo, Quinghua Li, Guohong Cao, Yigi Dai, Boleslaw Szymanski, Tom La Porta, “Social-Based Cooperative Caching in DTNs : A Contact duration Aware Approach”, IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, 2014

[23] H. Zhu, L. Fu, G. Xue, Y. Zhu, M. Li, and L.M. Ni, “Recognizing Exponential Inter-Contact Time in VANETs,” Proc. IEEE INFOCOM, 2010.

[24] L. Yin and G. Cao, “Supporting Cooperative Caching in Ad Hoc Networks,” IEEE Trans. Mobile Computing, vol. 5, no. 1, pp. 77-89, Jan. 2006.

[25] L. Yin and G. Cao. Supporting Cooperative Caching in Ad Hoc Networks. IEEE Trans. on Mobile Computing, 5(1):77– 89, 2006.

[26] J. Zhao, P. Zhang, G. Cao, and C. Das, "Cooperative Caching in Wireless P2P Networks: Design, Implementation, and Evaluation," IEEE Trans. Parallel & Distributed Systems, vol. 21, no. 2, pp. 229-241, Feb. 2010.

[27] Eiko Yoneki, Pan Hui, Shuyan Chan and Jon Crowcroft, "A Socio- Aware Overlay for Publish/Subscribe Communication in Delay Tolerant Networks," conference of MSWiM, 2007.

[28] Xu Liu ∗and Yuanzhu Chen Department of Computer Science Memorial University of Newfoundland "Report of A DTN Simulator - THE ONE" 2011.

[29] Wei Gao, Guo hong Cao, Arun Iyengar and Mudhakar Srivatsa, "Cooperative Caching for Efficient Data Access in Disruption Tolerant Networks," IEEE Transaction on Mobile computing, VOL. 13, NO. 3, MARCH 2014.

[30] Honglong Chen and Wei Lou, "Group Aware Cooperative Routing for Opportunistic Networks under Resource Constraints," conference of GLOBECOM, 2012.

[31] Feng Tian, Bo Liu, Jian Xiong, Lin Gui, "Efficient caching scheme for data access in disruption tolerant networks " 2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting in Ghent,17-19 June 2015.

APPENDIX

A.1 Tools and technology

This section describes the tools and technology used in our Thesis for calculation of the results. We have used the Opportunistic Network Environment Simulator (ONE SIMULATOR), NET-Beans and Eclipse as software package, for the performance evaluation and java technology to implementation of our project.

❖ Opportunistic Network Environment Simulator (ONE)

For the performance evolution in the DTN some proper tools is required and ONE Simulator is one such tool, the simulator is mainly used for nodal movement, modeling, inter node contacts, message handling and routing .one simulator has following important features.

- It can generate different types of mobility models, for example, random and map-based movements in the nodes of the network.
- We can code different types of routing algorithms for the transfer of the messages between the sender and receiver.
- We can see the movement of the nodes and the packets in the graphical interface; it also shows the summary of the ownership of packets at a particular time.

After the simulation, the collected results, and the analysis done by the post processing tools, in which we can visualize and, at the last, the reports are generated. The interactions of the elements are depicted in the figure 8. This simulator also enables the users to create their scenarios based on the mobility models that available and making use of the real world traces. The default trace that comes with this package is the map of Helsinki.

Once the configuration is s set, and start running the simulator, the GUI screen appears showing all the nodes in the network, figure 9. It is possible to customize the node as per our wish. The customizations could be altering the speed of the nodes, message creation intervals, buffer size, etc.

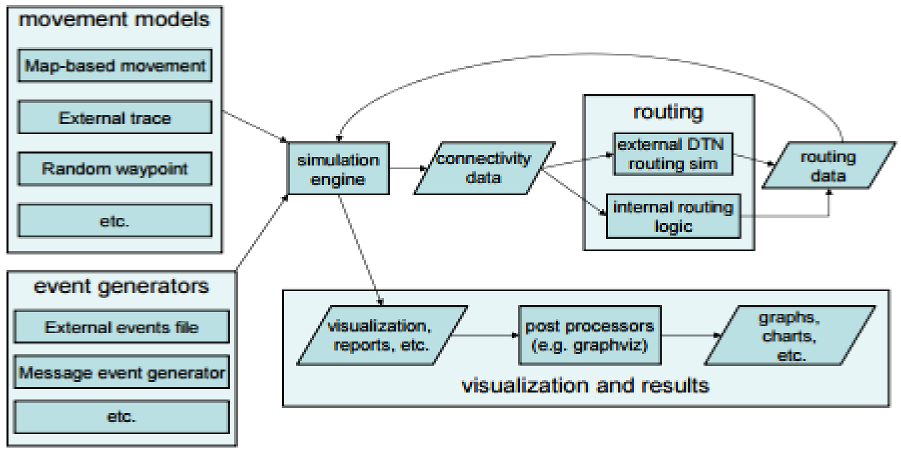

Figure A.1 Work flow of ONE Simulator [28]

Finally, after the simulation gets over then the log reports are collected which could be used for analyzing the problem. some loge file that ate logged are , the created messages report, delivered messages report, statistics of the simulation ,etc. it is also possible to customize the simulator to behave as per the needs , and get the things that is required to perform the analysis.

NET-Beans and Eclipse are the software development platform, implemented in java technology. Both are providing the reliable and flexible application architecture. It is easy to create application that is robust and complicated in eclipse and net-beans.