CRM: Cluster Based Replica Allocation over MANET

Rose Ann[1], Saritha S[2], Vipin Kuttamath[3]
Rajagiri School of Engineering and Technology [1,2] ,Huawei Technologies[3]

***Abstract*— Mobile ad hoc networks have shown a sporadic growth in recent years. The advancement in technology had leveraged its growth due to innumerable reasons. The comfort and easiness a wireless network can cater is the most important one. There had been a rapid growth of technologies and applications nurturing this development.The network partitioning that occur in course of mobile movements are the most aggressive feature. Link failures can isolate networks making data accessibility under question. Rapid and unpredictable topologies makes routing a herculean task,particulary in a system that lack a central authority or an arbitrator to negotiate routes on requirementBesides all these MANET are limited in action by storage and energy limitations.Hence, mobile hosts in MANETS take a selfish behavior to battle with the limitations of its construct. It tend to conserve its resources and solely utilize them for its own benefit. This in cooperative movement can hamper the overall goal of a mobile network. MANET requires full cooperation from all member hosts as to have a decent data accessibility. The nodes may however tend to adopt techniques to decrease its query delay rather than opting for the overall wellbeing. This paper work focusses on achieving a better performance in MANETs.**

***Index Terms*—About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

Technological advancement of this era has led to a sporadic emergence of wireless technologies. Mobile ad-hoc network (MANET) is one of the promising realm in it for research and development. The emergence of Personal Digital Assistants (PDAs), mobile Phones, handhelds, and wearable computers that enhance information processing and accessing capabilities with mobility played a vital role in improving quality of life and standards of living. Moreover, traditional home appliances, e.g. digital cameras, cooking ovens, washing machines, refrigerators, vacuum cleaners, and thermostats, with computing and communicating powers attached, extend the field to a fully pervasive computing environment.

A MANET is a collection of self-organizing wireless mobile nodes that can dynamically form a network on-the-fly to exchange information without using any fixed network infrastructure. This special features of MANET bring this technology great opportunities as well impose severe challenges. MANET is an autonomous system wherein mobile hosts are connected through wireless links. Each of the nodes in MANET acts as a router in the absence of any central server or a access point based physical infrastructure to arbitrate the routes.

MANET features a distributed kind of operation. Nodes constituting a MANET should fully collaborate among each other for its smooth functioning in the absence of central control for network operations. MANET often suffers from fluctuating links. The nature of high bit-error rates of wireless connection can be more profound in a MANET. One end-to-end path can be shared by several sessions. The channel over which the terminals communicate is subject to noise, fading, and interference, and has less bandwidth than a wired network. Mobile nodes are often limited by storage and energy requirements. Thus optimizations in computing and communicating mechanisms are of utmost importance.

Mobility in MANETs results in frequent network partitions as and when some nodes move away from an another node's communication range. This results in diminished rate of data accessibility. Mobility and resource constraints of mobile nodes results in performance degradation. Data Replication is one of the technique that can be used to address this issue. Nodes in a MANET are assumed to collaborate fully in terms of sharing their memory space. However, in reality some nodes may selfishly decide not to cooperate as to conserve their resources. The presence of such selfish nodes can reduce data accessibility and thereby hamper performance. Replication can simultaneously improve data accessibility and reduce query delay. This work focuses on addressing the selfish replica allocation problem in MANETs.[1]

II. RELATED WORK

The idea of data replication in MANETS were put forward by Kyu-Sun Shim ;SangKeun Lee ; Kun-Lung Wu, Jae-Ho Choi .The proposal involved methods to identify selfish nodes by use of a credit risk value. The method was similar to an economics subject wherein a nodes caliber to forward a packet was analyzed just like a bank checks the payability of a client taking a loan for repayment. A method similar to that was used in this thesis by using a bank that stores all the required information regarding the nodes in MANET. The packet flows was keenly analyzed using a network simulator NS2 achievd result indicated that the performance had substantially increased by applying a clustering technique in data replication.

III. REPLICA ALLOCATION IN MANETS

In a mobile ad hoc network mobile hosts access data items held by other mobile hosts. In a fully cooperative environment, each mobile host creates replicas of the data items, and maintains the replicas in its memory space. When a mobile host issues an access request to a data item, the request is successful if the mobile host holds the original or replica of the data item or at least one mobile host which is connected to the request issue host with a one-hop or multihop link holds the original/replica. Thus, first, the requested host checks whether or not it holds the original/replica of the target data item. If it does, the request succeeds on the spot. If it does not, it broadcasts the request of the target data item. Then, if it receives reply from other host which holds the original or replica of the target data item, the request is also successful. Otherwise, the request fails.

Replication can improve data accessibility and reduce query delay. The response time of a query can be substantially reduced if the query accesses a data item that has a locally stored replica. But it is not always practical to hold all of the accessed data in a local storage within the node due to its storage limitations. Hence we always look for a tradeoff between data accessibility and query delay. A node can only place a part of frequently accessed data in its local storage, or it may lead to all of the modes holding the same piece of information which in turn result in missing of data items with low access frequencies degrading the overall data accessibility.

We make some assumptions in the system model for replication. The assumptions are enumerated as below:

(i)Each mobile host in the network holds a unique host identifier. The set of mobile hosts is denoted by M={M1,M2,....Mm},where m is the total number of mobile hosts and each Mj is a host identifier ($1 \leq j \leq m$).

(ii)Each data item held by mobile host also has a unique data identifier. The set of all data items is denoted by D={D1,D2,....Dn},where n is the total number of mobile hosts and each Mj is a host identifier ($1 \leq j \leq n$).We also assume that the size of all data items are same and the data items held by a host is the original identifier of the host, that is M1 holds data D1,M2 holds D2 etc as their originals.

(iii)We specify a limit to the memory space held by each host, say C.

(iv)We also assume that the data items are not updated periodically. This helps us to rule out the inconsistency issues that me result from any stale data items or improper data updation or synchronization issues with data items.

(v)We also assume that the data items have their own access frequencies. Some data items are accessed more frequently than some others. This gives an indication on the need to have more replicas of data that are of more demand.

The fundamental activity to be done before any data replication is the analysis ofnetwork for data items and their access frequencies as mentioned in our assumptionsabove. Let Mi denote the number of mobile hosts connected to each other and Ni denotethe number of data items held by Mi mobile hosts. We then choose the optimal allocation practice by analytically finding the combination that gives high data accessibility. Following parameters are considered for this analysis.

(i) the access frequency from each mobile host to each data item,

(ii) the probability that each mobile host will participate in the network and will disappear from the network,

(iii) the probability that each two mobile hosts connected by a (one-hop) link will be disconnected, and

(iv) the probability that each two disconnected mobile hosts will be connected by a one- hop link.

Now we discuss some of the replica allocation techniques in use based on aaccess frequency table as illustrated below:

Data	Mobile host					
	M_1	M_2	M_3	M_4	M_5	M_6
D_1	0.65	0.25	0.17	0.22	0.31	0.24
D_2	0.44	0.62	0.41	0.40	0.42	0.46
D_3	0.35	0.44	0.50	0.25	0.45	0.37
D_4	0.31	0.15	0.10	0.60	0.09	0.10
D_5	0.51	0.41	0.43	0.38	0.71	0.20
D_6	0.08	0.07	0.05	0.15	0.20	0.62
D_7	0.38	0.32	0.37	0.33	0.40	0.32
D_8	0.22	0.33	0.21	0.23	0.24	0.17
D_9	0.18	0.16	0.19	0.17	0.24	0.21
D_{10}	0.09	0.08	0.06	0.11	0.12	0.09

Fig 1.1 Access Frequencies of Data Items for Mobile hosts

A. Static Access Frequency(SAF)

In this methodology, each mobile host allocates replicas of data items in descending order of the access frequencies. At the time of replica allocation, a mobile host may not be connected to another mobile host which has an original or a replica of a data item that this host should allocate. In this case, the memory space for this replica is retained free. The replica is created when a data access to the data item succeeds or when the mobile host connects with others.

B. Dynamic Access Frequency and Neighborhood(DAFN)

DAFN method aims to eradicate the pitfall of SAF method wherein replica duplication is visualized. This method achieves replica allocation in same way as that ofSAF on the basis of decreasing access frequencies. However it also further analyses theneighboring nodes in the network to determine any duplication. If any duplication isdiscovered then the mobile host with low frequency of data items get its replacement. Thus the duplication problem is substantially reduced by using DAFN.

C. Dynamic Connectivity based Grouping(DCG)

DCG method is an improvement over DAFN.[3]Here,instead of allocating replicas to a pair of neighborhood nodes, allocation is done within a larger group of nodes. The requirement of this methodology is that the group should be stable. It should sustain frequent changes in the network topology.DCG creates groups that are biconnected components in a network. This means that each group is a maximum partial graph which is connected if an arbitrary node in the graph is deleted. This makes the group highly stable as removal of a node from the group does not divide the group.DCG method is executed in every relocation period.

D. Detecting selfish nodes in MANET

Broadly there are two methods to detect selfishness, categorized as credit based schemes and reputation based schemes. Credit based scheme work by issuing credits to nodes for faithful their actions. The credit based scheme may be implemented using models like, the Packet Purse Model (PPM), Packet Trade Model (PTM), Secure Incentive Protocol (SIP) Sprite and auction based aodv protocol for mobile ad hoc networks with selfish nodes. Reputation based scheme operates by collectively detecting and declaring the misbehavior of a suspicious node. Such a declaration from the network nodes are then propagated throughout the network so that the misbehaving node will be removed from the rest of the network. Two models deployed under reputation based scheme are Path rater, Watchdog mechanism, CONFIDANT, CORE, token based approach, end to end acknowledgement schemes and reputation based algorithms.[5]

Nodes behavior selfishily in an attempt to reduce its query delay. However improving the overall data accessibility in a network is of utmost importance than the individual reduction in query delay.Malicious,faulty and misbehaving are terms interchangeably used in the domain of manets.Faulty nodes simply misbehave accidently. Selfish nodes exploit the routing protocol to their own advantage, mainly to nhance performance or conserve their resources. Selfish nodes show unwillingness to cooperate .Packet dropping is the main attack by selfish nodes, where most routing protocols have no mechanism to detect whether data packets have been forwarded. In the figure that depicts DCG implementation, it is possible that the node M3 may hold highly accessed data items like D3,D5 and D2 instead of data items D3,D9 and D2.This results in a scenario wherein other nodes in same group ,i.e. M1,M2 and M4 no longer being able to get access to D9.This in turn reflects in its degraded data accessibility rate since M1,M2 and M4 cannot fully leverage the memory space of M3.This case is of Fully selfish behavior.

Also there can be a partial selfish behavior wherein a node uses a part of its local storage for its own benefit and a part for the overall network performance. An example taken may be node M4 in DCG figure. Here instead of holding replicas of D4,D6 and D10,Node M4 may selfishly decide to hold D4,D6 and D2 for its partial selfish behaviour.This also results in performance degradation.

Selfish nodes may be detected in a real scenario by calculating a parameter called as credit risk. The value may be calculated as ,

Credit Risk=expected risk/expected value

Each node makes a measurement of this score for each of the nodes connected to it by a link. This gives an estimation of the degree of selfishness for all of the connected nodes in the network. Selfish features belongs to two main categories ,namely node specific feature and query specific feature. A partially selfish node utilizes only a part of its ocal storage for overall goodness of network. Thus it replicates only some of the data items it is supposed to replicate. Thus the number of data items replicated or the size of the local memory utilized for shared data items gives the node specific feature or the expected value of a node. The size of Nk's shared memory space, denoted by SSik and number of Nk's shared data items NDik gives the expected value. The values are however estimated values as a node never lets another node know the exact replicas it shares nor the memory space allotted for the common goal. A query specific feature is given by the ratio of selfishness alarm of Nk on Ni, denoted by Pik.This is defined as the ratio of Ni's request not served by expected node N k due to its selfishness characteristics. This characterizes the expected risk of a node. The risk value of a node increases as Pik goes higher and higher.

$$CR_i^k = \frac{P_i^k}{\alpha * SS_i^k + (1-\alpha) * ND_i^k}, \text{ where } 0 \leq \alpha \leq 1$$

α is used to adjust the relative importance of SSik and NDik..

IV. CLUSTER BASED REPLICA ALLOCATION IN MANETS

Replications allow better data sharing. The performance and data accessibility of MANET's can further be improved by reduction of communication and traffic overheads. An improvement over the conventional practices is to regulate the traffic flow within small groups of nodes rather than spreading the traffic over a large area. This can substantially reduce the overheads and thereby improve network performance. Hence our task is to integrate replication and clustering schemes in areas populated with selfish nodes. Replication can hence be done within small groups and thereby accessibility as well as performance can be improved. DSR supports packet salvaging .Salvaging logic differ from time to time and is defined by the DSR agent.[4]

The SCF tree based replica allocation is based on human friendship management wherein each one manages friendship at his/her own discretion. It does not require any lengthy discussions or negotiations to maintain the same. The decision is solely dependent on one's own discretion. This means any novel replication schemes based on this behavior can reduce traffic overhead and has improved data accessibility. We reduce the graph containing selfish and non-selfish nodes to a partial sub graph which is a component of graph. As the SCF tree consist only of non-selfish nodes and hence we measure the degree of selfishness to create the tree. We set a faulty rating threshold and a good rating threshold. The degree of goodness or selfishness can hence be detected. We implement a mechanism to monitor the energy values, queue length and number of forwarding's done by a node periodically.The packet flows are analyzed in the network and as and when it arrives a node, they are cached.

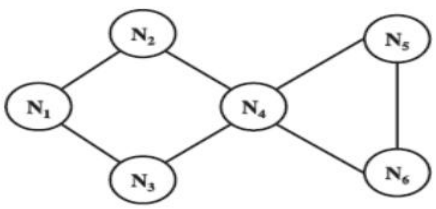


Fig .2 Sample Topology

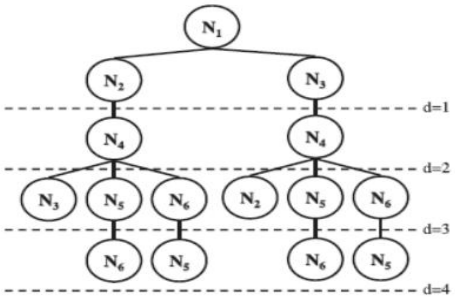


Fig.3 SCF Tree of N1

After building SCF tree, a node allocates replica at every relocation period. Each node asks non selfish nodes to hold replica. Replica allocation is done independently at its own discretion. By observing behaviors ,it updates expected risk and expected values at every relocation period. Memory area of every node is divided into selfish are and public area. Selfish area is used to save replicas of local interest whereas public area is used to hold data for others, that is to improve data accessibility. Data is replicated to the closer nodes after cluster formation in the order of decreasing access frequencies. The priority for allocating replicas are on the breadth first search of the SCF tree.

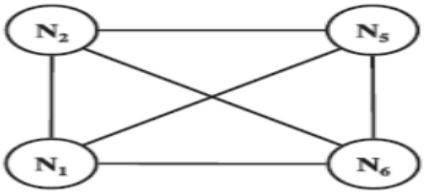


Fig .4 Graph of Non Selfish Nodes

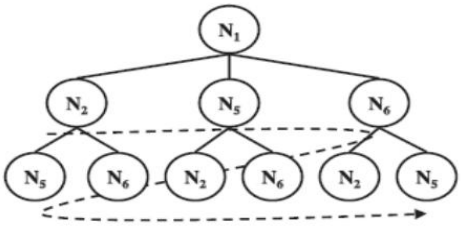


Fig .5 SCF Tree of N1

SCF tree provides good data accessibility with low communication cost in the presence of selfish nodes.

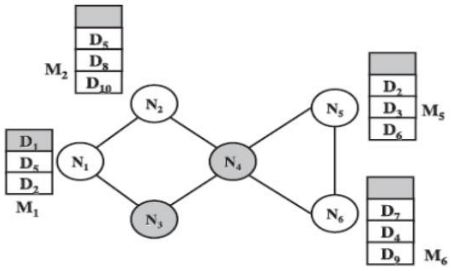


Fig .6 Expected Replica Allocation result

N_i	N_k					
	N_1	N_2	N_3	N_4	N_5	N_6
N_1	·	0.30	0.85	0.80	0.45	0.22
N_2	0.40	·	0.80	0.90	0.30	0.50
N_3	0.25	0.35	·	0.75	0.65	0.75
N_4	0.45	0.44	0.51	·	0.23	0.37
N_5	0.30	0.60	0.85	0.40	·	0.21
N_6	0.40	0.50	0.90	0.52	0.30	·

Fig.7 nCRik for each node

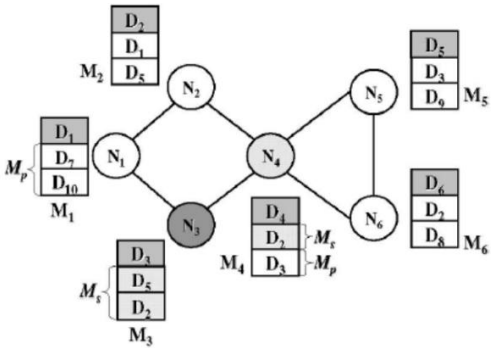


Fig.8 Final Replica Allocation Result

V. SIMULATION AND RESULT

The simulator we have chosen for our simulation is NS2. Our system model comprises of inputting a set of malicious, selfish and faulty nodes to the network structure . Malicious and selfish nodes differ in the way they operates. Malicious node forwards route reply and route request packets and drop all other packets. Selfish nodes drop route request and route reply packets. Faulty nodes on the other hand, drop all packets. Faulty nodes are penalized for their behavior by not forwarding packets originated from them. Normal nodes drop their packets. Normal forwarding patterns are followed in case of selfish or malicious node. Nodes do not forward any packets if network is congested. Also if faulty nodes are contained in route reply, routes are discarded.Table 1 gives a striking comparison of the packet delivery ratio,Delay and throughput achieved in the replication scheme using DSR.

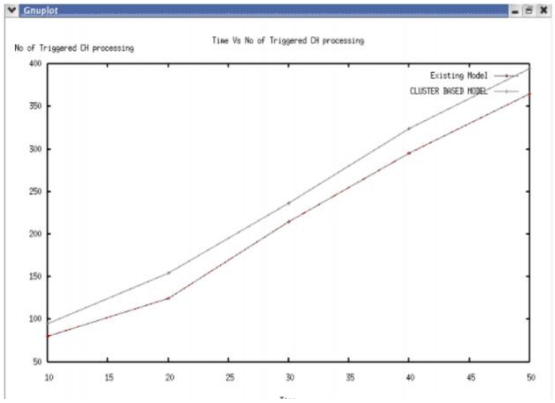
	PDR	Delay	Throughput
Existing System	99.05	0.211	1760.96
Cluster based replication	99.15	0.205	1492.08

Table 1. Performance Evaluation

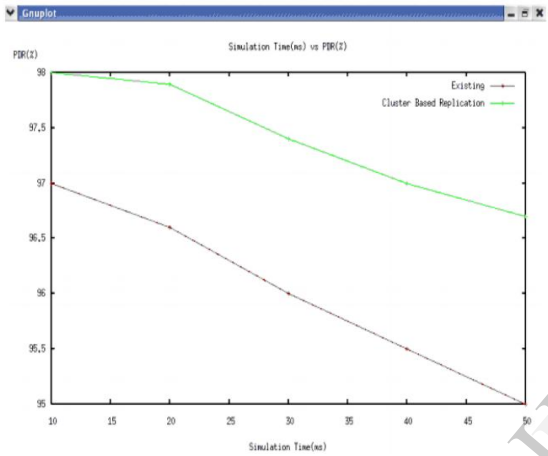
```
Packet Delivery Ratio     : 99.0476190476191
Average End2End Delay     : 0.204828355605769
Average Number of Hops    : 1.11538461538462
Control Packet Overhead   : 393
Throughput                : 1760.96931587459
Data Packets Sent         : 105
Data Packets Received     : 104
Simulation Endtime        : 2.362335313
Total Delivery Time       : 21.302148983
Total Number of Hops      : 116
Dropped Reply Messages    : 2
Maximum Number of Hops    : 4
Minimum Number of Hops    : 1
```

Table 2 Output File

TABLE 2 gives the parameters extracted from the trace file generated in DSR simulation.



Graph.1 Cluster head processing comparison



Graph.2 Packet Delivery Ratio

VI. CONCLUSION

The thesis work on CRM,cluster based replica allocation over a mobile ad hoc network witnessed improvement in performance in terms data accessibility and improved delay characteristics.The study focused on simulating a network environment supported by OCEAN protocol of DSR enhancement.The method was found to have improved performance than using conventional DSR.Further ,it was discovered that data replication techniques are an efficient means to provide guaranteed service to any critical applications in industry.The flexibility and widespread use of MANET in recent years has evolved due to the technological advancement in this era. The primary step before duplication of data is identification of selfish node.Mobile nodes have random motion and mobility of mobile nodes often causes link breakages.this can cause packet drops which could be misunderstood as the scenario with selfish nodes.Thus false alarms generated in a network is an area of reasrch to be continued in the light of selfish misbehaviours.

REFERENCES

[1] David Oliver Jorg,"Performance comparison of MANET Routing Protocols In Different Network Sizes",Computer Networks and Distributed systems,2003.

[2] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in Proc. of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom), August 2000.J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," IEEE Trans. Electron Devices, vol. ED-11, pp. 34-39, Jan. 1959.

[3] 23. Jae-Ho Choi,Kyu-Sun Shim,SangKeun Lee and Kun-Lung Wu,"Handling Selfishness in Replica Allocation over a Mobile Ad Hoc Network",IEEE transactions,Vol.11 February 2012

[4] David B.Johnson,David A. Maltz,JoshBroch,"DSR:The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks",Ad Hoc Networking,2001

[5] Rose Ann and Saritha S:A survey on evaluation schemes for routing misbehaviors in MANET In: International Journal of Computer Application,Volume 78,No.11,September 2013.