

A Survey on Query Optimization and Data Replication Techniques in Wireless Adhoc Networks

Vijay Kumar S¹

Assistant Professor, Department of Information Science and Engineering, BNM Institute of Technology,
Bengaluru, India¹

ABSTRACT: Maintaining Data available all time and Link Failure are the two major issues to be addressed in the wireless Adhoc network. Due to the link failures the network partition takes place and the requested data is not transferred to the destination. Due to this problem we need to set up the path once again to transfer the requested data to the destination node or sink node. This increases the time taken for the whole process and the efficiency is reduced by a greatest extent. To address this issue many schemes are proposed. These schemes deal how to reduce the delay induced in the querying process and increasing the data availability at the same time. In this paper we compare the different schemes altogether so that the basis of the work can be understood by the new researchers. There are data replication schemes that are defined for increasing the availability but the main problem in this approach is that the mobile nodes don't have that much capacity to store and buffer is over flown most of the time. The availability of data is inversely proportional to the time required for the data access. As the data availability is more the access time reduces.

KEYWORDS: Replication, data availability, wireless adhoc network, data access and Link Failure

I. INTRODUCTION

In mobile ad-hoc networks, frequent network partitioning and the failure of mobile nodes are due to exhaustion of their battery power. Most of the time mobile nodes use the battery for path setting and searching for a data (data availability). Whenever there is an increase in network size and node mobility it causes the performance of data access to degrade. To deal with these issues, a number of data replication protocols have been proposed in the recent years. In ad hoc networks, there are many applications in which mobile hosts access data items held by other hosts. Here, if a network partition occurs due to the movement of mobile hosts, data in two separated networks becomes inaccessible to each other. Thus, data accessibility in ad hoc networks becomes lower than that in the conventional fixed networks. A key solution is to replicate data items on mobile hosts that are not the owners of the data items, of course it increases storage overhead, but in this paper we are concentrated on data availability and data transfer over link failure condition.

The figure 1 gives the overview of the network partition and the data availability.

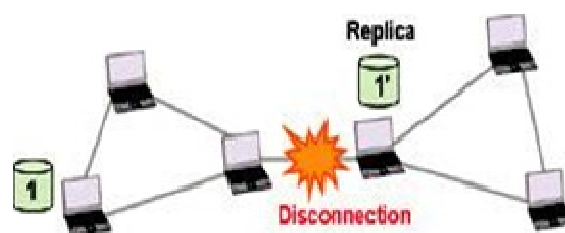


Figure 1: Network partition and data replication

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

In a real environment, since it is more likely that data items are updated, mobile hosts may access old replicas that have been updated. Accesses to old replicas are invalid and cause useless data accesses and rollbacks. We assumed an environment where data items are updated and proposed replica allocation methods which consider read/write patterns or data update intervals. [7-17].

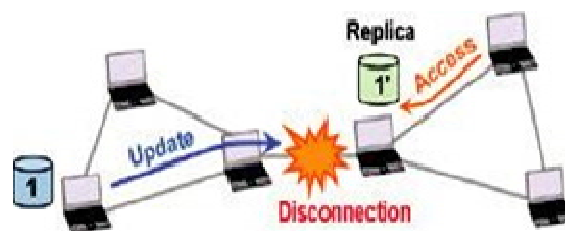


Figure 2: Data update and access.

In mobile adhoc networks, since the communication bandwidth is limited, it is important to acquire only necessary data items to reduce data traffic. A possible and promising solution is that each mobile node retrieves data items using a top-k query, in which data items are ordered by the score of a particular attribute and the mobile node that retrieves data items (the query-issuing node) acquires ones with the k highest scores (top-k result). In Figure 3, we show an application example of a rescue effort at a disaster site where the communication infrastructure, e.g. Internet, has been broken down. Each rescuer can search for several victims with serious injuries to evacuate them with a limited number of ambulances. The top results can also be applicable for the military application because the best possible efficient solution is taken into consideration. Here, if each rescuer transmits the information on the two most injured victims in his/her responsible region, the query-issuing rescuer acquires information on more victims than necessary. In addition, if the radio link between portable computers of rescuers in the bottom-right and top-left corners is disconnected, the query-issuing rescuer cannot acquire information on victims managed by the rescuer in the bottom-right corner.

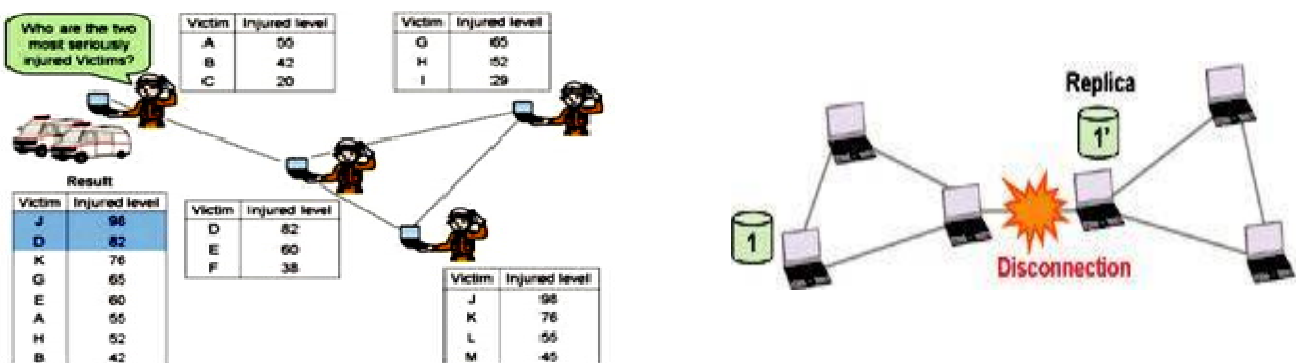


Figure 3: Query optimization

In this paper will take the algorithms defined for the bridging of the delay and data availability.

In this paper will take the algorithms defined for the bridging of the delay and data availability.

In ad-hoc networks, link failures are generally due to node mobility and link hysteresis. To emulate link failure behavior in the indoor environment, intermediate nodes along the paths are switched on and off once every 10 seconds

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

during data transmission among nodes. In our experiments, every scenario consists of an alternate path to continue data transmission in case of link failure events.

In ad hoc networks, since it is more likely that battery capacities of mobile hosts are limited, mobile hosts at the center of the network and holding data items (replicas) frequently accessed by other hosts need to transmit them many times and consume much power than other hosts (Fig. 8). In this case, mobile hosts cannot access data items held by the mobile hosts that exhaust their batteries and leave the network, and thus, data accessibility becomes lower. There are replica allocation methods for not only improving data accessibility but also balancing the power consumption among mobile hosts. In these methods, each mobile host replicates data items frequently accessed by itself and its nearby hosts. As a result, these methods can balance the numbers of data accesses performed on data items held by each host. In addition, we have proposed the following data access methods:

- To prevent mobile hosts from exhausting their batteries, each mobile host selects the paths on which mobile hosts have much remaining battery power and uses it for data transmission.
- To maintain data availability and prolong the lifetime of mobile hosts for reducing data traffic, each mobile host collects multiple requests for data items and transmits the requested data items by multicast.

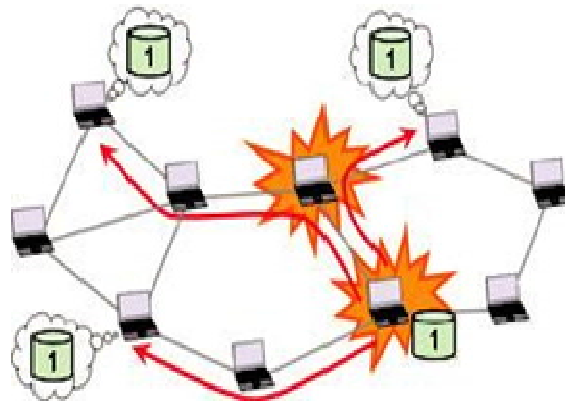


Figure 4: Decrease of remaining amount of battery by skewed data accesses.

II. REPLICATION OF DATA

A. Problem of replication of data

Mobile hosts run on limited energy source (i.e. battery power) and communicate with each other via wireless links of limited bandwidth. Hence, it is necessary to incur low communication cost during the transfer of information from one mobile host to another. In an ad hoc environment, ensuring reliability of obtained information is a major problem due to frequent partitioning of the network as a result of the mobility of servers. It becomes absolutely necessary to improve the processing time for transactions while respecting battery power considerations and ensuring the reliability of information. Data Replication in Ad hoc mobile databases is gaining importance due to the need for an effective replication scheme and various complications involved in replicating data for ad hoc systems. Efficient data replication among the energy sufficient hosts could result in providing better and reliable sources to database queries originating from mobile hosts. Available techniques give solutions for replication in mobile ad hoc networks with compromises on time and consistency respectively.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

B. Algorithms defined for data replication

DAFN is a scheme that is used for checking the access frequency for the data item by both the nodes that are neighbor's to each other. Consider an example as shown in the figure 4 where there are two nodes n1 and n2 that are trying to access data item d1, d2, d3, d4. We need to take the access probability of the nodes on the considered data item and check whether the required data item is already present in the neighbor or not. The main intention of this algorithm is that the data should be present in the single hop. The neighbors should not contain the same data so if we see the table below we can understand the way the access probability is taken into consideration.

Data	Node Number	
	N1	N2
D1	0.6	0.4
D2	0.2	0.4
D3	0.1	0.2
D4	0.1	0.2

TABLE 1: DATA ACCESS PROBABILITY

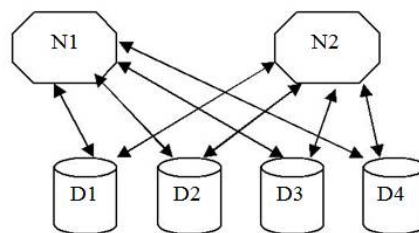


Figure 5: DAFN system

In the first pass the nodes replicate the data depending upon the interest they have but if you take the consideration of the later step the data access probability is taken into consideration and the probability that is high is taken by the node and other is defined for another one. So in the case of table 1 node 1 will have d1 and d3 and node 2 will have d2 and d4 respectively. In Static Access Frequency (SAF) algorithm, allocates replicas of data in descending order of the access frequencies from each node (within the limit of its own memory space). Since the nodes do not need to exchange any information with each other, SAF method allocates replicas with low tra c but when hosts have similar access characteristics, it excesses duplicates data item among them. Whereas Dynamic Access Frequency and Neighborhood (DAFN), eliminates the replica duplication among directly connected nodes. If there is a replica duplication of data item between two neighboring nodes, the node with the lower access frequency (relative to the data item) changes the replica to a replica of di erent data item. It is accomplished by a repeated procedure of the breadth first search (from the host with the lowest id) in each set of nodes which are connected to each other. Dynamic Connectivity based Grouping (DCG), shares replicas in larger groups of mobile nodes than the DAFN. The groups used for replica allocation in MANETs should be stable, that is e.g., the group is not easily partitioned due to changes of network topology. DCG method uses biconnected components to create groups of nodes in MANET. A biconnected

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

component is defined here as a maximal partial graph G_0 of G which cannot be divided by deleting any vertex $v \in G$. DCG protocol is executed at every relocation period and initially each node broadcasts its id and information on access frequencies to data items, which allows hosts to recognize the connected nodes. In each partition, the node with the lowest id executes an algorithm to find biconnected components with the known network topology at the moment – each biconnected component is put into a group. Hosts that belong to more than one biconnected component can only belong to one, e.g. to one which was found first. Next, the host with the lowest id in each group (coordinator of the group) calculates an access frequency of the group to each data item as a summation of access frequencies of all nodes in the group to the data. Then, the coordinator determines replicas allocation in the group in the order of the access frequencies of the group – each replica is allocated at the host whose access to the corresponding data item is the highest (allowing hosts free memory space).[15]

III. GREEDY SCHEME

Consider if there are two nodes n_1 and n_2 and the data they access are d_1, d_2, d_3, d_4 out of which they access the d_1 data more and this is the most recently and frequently accessed data then the replication of that data is done first. Consider the fact that is we take the value of access frequency of the node n_1 accessing the data d_1 and divide it with the size of the data item d_1 then if the value is maximum of all the permutations then if it matches with the data item present in n_1 then it is discarded or the data is put forth in the node n_1 base.

The main disadvantage in this case is that it does not take into consideration the cooperation between the neighboring nodes and due to this the performance of this algorithm is limited.

IV. ONE TO ONE OPTIMIZATION SCHEME

A. Basis

In this optimization scheme that was included in the paper [22] the nodes that are neighbors to each other cooperate to make the replication possible. The combined access frequency of the two nodes is taken into consideration that is defined as CAF_{ij}^k . For it we will take the access frequency of the access of data item d_k by the node i and node j into consideration. We will take the link failure probability also into consideration. Consider the figure 5, in the figure we can see that the node 1 and node 2 are connected to each other and node 1 is having the data 1 and the equation of the CAF_{ij}^k will be given as

$$CAF_{ij}^k = (a_{ik} + a_{jk} \times (1 - f_{ij}))/s_i \quad (1)$$

$$CAF_{ji}^k = (a_{jk} + a_{ik} \times (1 - f_{ij}))/s_i \quad (2)$$

Later we need to assign the weighted value to the CAF such that we can define the levels of priority of the data access on the data item. The weight function is defined by w and multiplied by CAF to get the value.

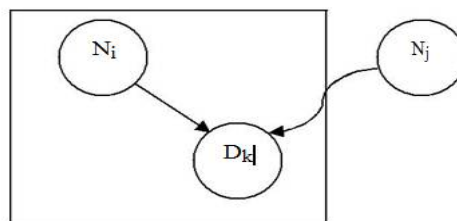


Figure 6: OTOO

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

The value function is designed so that 1) it considers the access frequency from a neighboring node to improve data availability; 2) it considers the data size. If other criteria are the same, the data item with smaller size is given higher priority for replicating because this can improve the performance while reducing memory space; 3) it gives high priority to local data access, and hence the interested data should be replicated locally to improve data availability and reduce query delay; 4) it considers the impact of data availability from the neighboring node and link quality. Thus, if the links between two neighboring nodes are stable, they can have more cooperation in data replication.[13]

V. RELIABLE NEIGHBOR SCHEME

Consider if we are taking the OTOO scheme the node will take the value of its own accessibility for the data item so the reliable scheme will take the threshold Tr and the value of link failure probability f_{ij} is calculated so the equation defined is $(1-f_{ij}) > Tr$.

Consider that $nb(i)$ is the set of neighbor nodes for $N(i)$ then in that case the reliable neighbor replicates the data for neighbor nodes until the memory is full. [12]

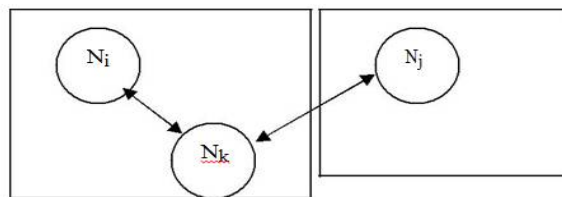


Figure 7: RN Scheme

VI. RELIABLE GROUP SCHEME

OTOO only considers one neighboring node when making data replication decisions. RN further considers all one-hop neighbors. However, the cooperation's in both OTOO and RN are not fully exploited. To further increase the degree of cooperation, we propose the reliable grouping (RG) scheme which shares Replicas in large and reliable groups of nodes, whereas OTOO and RN only share replicas among neighboring nodes. The basic idea of the RG scheme is that it always picks the most suitable data items to replicate on the most suitable nodes in the group to maximize the data availability and minimize the data access delay within the group. The RG scheme can reduce the number of hops that the data need to be transferred to serve the query. [10]

VII. CONCLUSION

This paper will serve as a basic reference to understand the work carried out by the researchers on reduction of delay and increasing the availability of data between the nodes. In future we can define efficient and enhanced methodologies to address these issues by analyzing the existing schemes discussed in this paper.

REFERENCES

- [1]Hara, T., Loh, Y.-H., and Nishio, S.: Replica Allocation in Ad Hoc Networks Considering the Stability of Radio Links, IPSJ Journal, Vol. 44, No. 9, pp. 2308-2319 (Sept. 2003).
- [2]Hara, T., Murakami, N., and Nishio, S.: Replica Allocation for Correlated Data Items in Ad-Hoc Sensor Networks, ACM SIGMOD Record, Vol. 33, No. 1, pp. 38-43 (Mar. 2004).
- [3]Hara, T.: Effective Replica Allocation in Ad Hoc Networks for Improving Data Accessibility, Proc. of IEEE Infocom 2001, Vol.3, pp.1586-1576 (Apr. 2001).
- [4]Hara, T., Loh, Y.-H., and Nishio, S.: Data Replication Methods Based on the Stability of Radio Links in Ad Hoc Networks, Proc. of Int'l Workshop on Mobility in Databases and Distributed Systems (MDDS 2003), pp. 969-973 (Sept. 2003).

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

- [5]Hara, T.: Data Replication Issues in Mobile Ad Hoc Networks, Proc. of Int'l Workshop on P2P Data Management, Security and Trust (PDMST 2005), pp. 753-757 (Aug. 2005).
- [6]Hayashi, H., Hara, T., and Nishio, S.: A Replica Allocation Method Adapting to Topology Changes in Ad Hoc Networks, Proc. of Int'l Conf. on Database and Expert Systems Applications (DEXA 2005), pp. 868-878 (Aug. 2005).
- [7]Hara, T.: Replica Allocation Methods in Ad Hoc Networks with Data Update, ACM-Kluwer Journal on Mobile Networks and Applications (MONET), Vol. 8, No. 4, pp. 343-354 (Aug. 2003).
- [8]Hayashi, H., Hara, T., and Nishio, S.: Updated Data Dissemination Methods for Updating Old Replicas in Ad Hoc Networks, ACM/Springer Personal and Ubiquitous Computing, Vol. 9, No. 5, pp. 273-283 (Sept. 2005).
- [9]Hara, T.: Replica Location Management for Data Sharing in Mobile Ad Hoc Networks, Journal of Interconnection Networks (JOIN), Vol. 7, No. 1, pp. 75-89 (Mar. 2006).
- [10]Vijay Kumar S, PalivelaHemant , Pinaki Barman “Skyline Query Processing In Spatial Databases Using Data Mining Based Techniques”, Second International Conference On computation Of Power, Energy, Information And Communication, ICCPEIC- 2013 , 17-18 April 2013, ISBN: 978-1-4673-6406-5/13/&31.00
- [11]HemantPalivela, Prof. Yogish H K, Vijay Kumar S, KalpanaPatil “A Survey on Mining Algorithms for Breast Cancer Related Data” IEEE Sponsored International Conference On Information Communication & Embedded Systems "ICICES 2013", 21st -23rd February 2013, ISBN : 978-93-82-880-03-5.
- [12]HemantPalivela, Prof. Yogish H K, Vijay Kumar S, KalpanaPatil “A Study Of Mining Algorithms For Finding Accurate Results And Marking Irregularities In Software Fault Prediction” , IEEE Sponsored International Conference On Information Communication &Embedded Systems "ICICES 2013", 21-23 February 2013, ISBN : 978-93-82-880-03-5.
- [13]Vijay Kumar S, Ravi P , Vijay Kumar P “Hand Drawn Routers for Robot pilot age on Android Platform” , National Conference on Emerging Trends in Computer Science and Engineering conducted at NCET, Bengaluru during April 24th - 25th 2015.
- [14]HemantPalivela, Vijay Kumar S, KalpanaPatil, PrashanthG “Voting Based Approach To Reduce Impact Of Statistic-Poisoning Attacks In CIDS” National Conference On Advanced IT, Engineering And Management, 6th-7th September 2012, SAC AIM-2012, St. Aloysious College, Mangalore, Karnataka
- [15]Mrs.Sharada K A , HemantPalivela, , Prashanth, Vijay KumarS “A Model Proposed For Reducing The False Positive Alarm Rate Using The Feature Of Event Correlation “International Journal Of Advanced Research In Computer Science And Software Engineering (IJARCSSE), ISSN: 2277-128X Vol. 2, Issue 8, August 2012.