Data Replication in Current Generation Computing Environment

[1]P J Kumar
VIT University, Vellore, India

[2]P Ilango
VIT University, Vellore, India

Abstract

MANET, VANET, CLOUD and IoT are some of the recent computing models or network architecture that has evolved in the recent times. The characteristics of devices and communication links of these types of models vary largely leading to high degree of heterogeneity. The applications designed for such a heterogeneous type of network faces challenges such as frequent network partition, communication link failure, data node crash and variations in the performance of network links etc. Replication of data is a technique which increases data availability in case of occurrence of problems as mentioned above. Though the name replication sounds familiar, its implementation to different networks demands a custom methodology according to the type of network. In this article we study various replication methods used in different network architecture or computing model. We identify the challenges in replicating data in diverse environment and list the scope for possible extensions in future to improve the performance of replication.

Keywords: *Data Replication, VANET, High data or service availability, MANET, CLOUD, P2P, Energy Efficiency*

1. Introduction

Node mobility causes frequent network partition in networks such as mobile ad hoc network and Vehicular ad hoc network. A node may not be able to access the data maintained by the other node during network partition. Data is replicated on several other nodes to improve accessibility [7]. Similar to data replication web services can be replicated on several web servers to increase service availability of in case of server crash [5][8]. While increasing data availability through replication, it must be ensured that the overhead associated with replication does not increase beyond a threshold. The varying features of nodes in MANET and VANET such as limited memory, battery power, constrained bandwidth , frequent link disconnection demands different methodology to replicate data [4][6]. Unlike the computing devices available as part of wired the network, devices in mobile ad hoc network has limited memory in it.[4][6]. Node cooperation is an important characteristic of mobile devices to leverage data availability. Nodes with better cooperation in extending their memory for replicated data helps to improve data availability. But certain nodes extend only less cooperation to share memory for the benefit of others. Recent works handle these characteristics of nodes called selfishness [6]. In contrast with mobile networks, Cloud offers large storage for users to organize data. [1]. While supporting large volume of data for users cloud performs replication to ensure QoS guarantees for applications running in the cloud. [10]. Vehicular Ad hoc network (VANET) consists of vehicles moving in a particular direction with a particular velocity. A node in VANET has more computing capacity, battery power than the nodes in MANET. A group of vehicles moving within a boundary limit can communicate and share data available with every other node and communicate with the Internet through satellite. VANET is used to forecast the danger ahead in road, secure the passengers by providing health services by interacting among vehicles in the network. Changes in the velocity of the vehicles causes frequent disconnection between vehicles and causes unavailability of data among vehicles. It is important to identify vehicles on which the data can be replicated to increase availability of data during network partition. Several algorithms has been proposed in the literature to replicate data in MANET, VANET [4][6] and CLOUD [1][2][3][10]environment. In this article we perform a survey on various replication algorithms proposed in the literature and perform a comparative analysis as shown in Table:1 and also provide scope for improvement.

2. Various Approaches for Replication

2.1 Maximising P2P file access in MANET

The work in [11] performs replication to increase file availability among peer nodes in MANET. The approach considers the meeting frequency of the

nodes as a parameter to choose nodes to replicate data on it. Since a node that has higher meeting frequency will be communicated by several other nodes in the network. It also identifies policies to allocate resources for nodes based on the measured attribute.

The work [11] has simulated the proposed algorithm PCS in NS2 and GINI. The content of the file is not modified during the access. It remains static. Replication with file update is considered as part of the future work which will consider the update of files during access.

2.2 Replication in VANET

A group of moving vehicles forms a network to share information and data among other vehicles through access points installed along road side. An example network is illustrated in figure 1.The communication link between vehicles breaks often due to high velocity of nodes. Data are replicated in many nodes to provide continuous availability to vehicles in presence of frequent link failure. Several tasks are performed such as node position calculation, accessing data from other vehicles and transferring the packets to the destination vehicles. RFID tags are used in the vehicle nodes and RFID readers are used in the Road Side Units to gather information among vehicles. This method eliminates installation of access points on the road side. Bloom filters are used to find the nodes on where replica can be placed. It is better than the GPS based mechanism to fid nodes in a sparse distribution of vehicles.

The work in [13] proposed a protocol called as Replication Aware Data dissemination to place replica on nodes. The proposed approach has been tested on different network configurations with various performance measures.

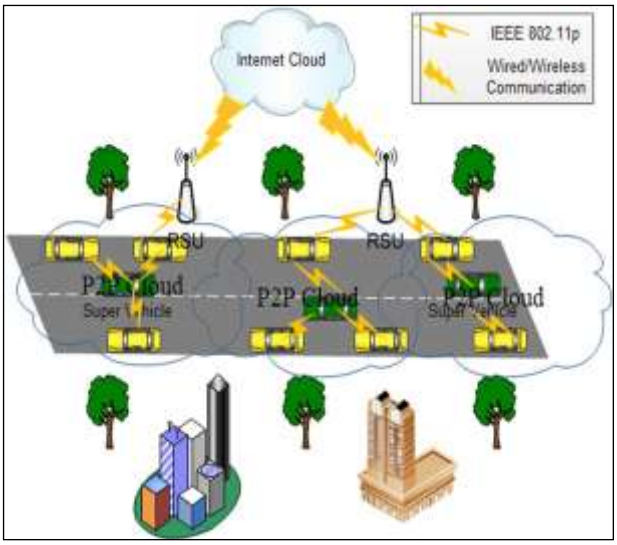

Figure 1: Replication in Vehicular Ad hoc Network -Excerpt from [13]

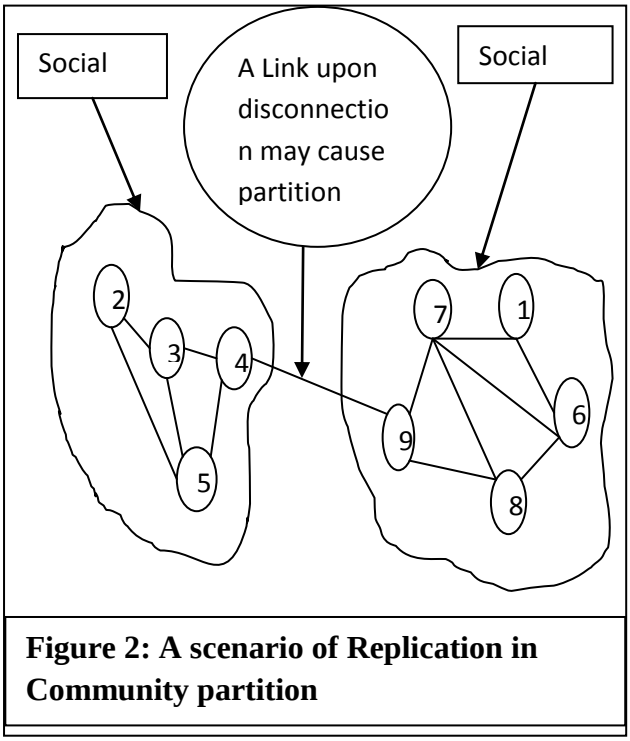

Figure 2: A scenario of Replication in Community partition

2.3 Replication in ASNET

An ad hoc social network (ASNET) has gained more popularity due to the widespread usage of social media applications. It is a combination of MANET and ASNET. In recent time, there has been a tremendous increase in the number of users in the social media and the related applications such as Facebook and WhatsApp. As shown in the figure 2, the members of a particular social media can access the information from the other users of

the group. A user can access the information from the other social groups, provided, there exist a communication link between two groups. As shown in the diagram the link from node 4 and 9 enables node 2 to access data from node 6 though they are from different community. In case of link breakage between the node 4 and node 9, the node from one community cannot access the data contained by the other community. Replication improves data availability in ASNET during a communication link failure. It is important to select nodes and replicate data to them so that the other nodes can have access to data during network partition.

The work [12] proposed (community partition aware Replication for ASNET) to replicate data on several nodes to increase availability.

2.4 Dynamic data replication strategy in cloud

Replication is performed in cloud environment where node failure occurs frequently. The number of replicas placed is an important criterion for a better performance. The proposed approach determines the optimal number of replicas and the place where the it can be replicated.

The work in [9] performs an analysis and models the relationship between the number of replicas and system availability. It also evaluates the popular data and determines the number of replicas to be placed. It proposes a dynamic replication strategy. In addition to the proven benefits of the algorithm, minimizing waiting time, reducing data access time and improving data availability further are considered as an extension work.

2.5 Energy aware Data Replication in Cloud Computing Data Centres.

Cloud offers various services as pay per use basis to its users. The applications which makes use of the cloud services accesses data from data centres which are far away from the applications. It is better to place data near to the application which accesses it. This reduces the energy required to transfer the data from data centres to the application and also it reduces the access time, network bandwidth and increases the quality of service by reducing the communication delay.

The work [14] has tested the proposed energy efficient replication scheme using Green Cloud simulator which is energy aware simulation for cloud.

2.6 Load balancing in cloud using replication

Load balancing in cloud is performed among applications that are executed on nodes in the cluster. An execution request for an application is replicated on several queues maintained by servers. The server that handles the request first executes the application in it. It also sends a notification to the other servers that a has request for the application to execute. The request is removed from the queue as soon as a notification is received. The overall waiting time of the request in a particular server is reduced to a large extent.

The work [15] has shown that the proposed system reduces the queuing time significantly even with less number of replicas under heavy system loads.

3. Comparative study

S. No	Name of the Replication methodology	Computing model	Objectives	Scope for Extension
1	Priority Competition split replication protocol	Peer to Peer file access in Mobile ad hoc network	An pproach to improve the file access and its availability among peers in MANET	Extending the work to consider the updates performed with the file and the propagation of updates to peers.
2.	Replication Aware Data dissemination using bloom filter to select candidate vehicle for replication	Resource sharing and communication that takes place between vehicles which are part of VANET	To increase availability of data among vehicles in VANET either in the sparse region or in the dense region.	Addressing security measures while replicating data on vehicle nodes in VANET.

3.	Community partition aware Replication for ASNET	Group of people sharing and communicating data or information through MANET	To improve effective data sharing and information propagation among social network peers.	To increase data accessibility among peers in social ad hoc network
4.	Dynamic Replication Strategy	Cloud Computing	Maintaining a balance between the numbers of replica allocated and the place where a replica is placed is the main objective of this approach.	Reduction in waiting time of the user and possibilities to increase data availability further
5.	Energy Efficient Replication	Cloud Computing	Reducing the energy consumption by replicating data in a nearby node.	Identifying further measures to reduce energy consumption
6	Load balancing using replication	Cloud Computing	Decreasing the queuing delay faced by the applications while executing it on cloud servers.	Increasing the performance of the system in order to reduce the queueing delay further.

Table 1: Comparison of various Replication methodologies

4. Conclusion

Replication has been a popular technique to increase data availability or service availability in presence of network partition, link failure, server crash etc. Though the term replication looks familiar since a long time back, the demand for a custom replication methodology evolves along with the evolution of computing or network architecture. In this paper we have analysed various replication techniques adopted in different computing paradigm such as MANET, VANET and Cloud. We have studied the objectives and functionalities of various approaches for replication and suggested possible extensions as a future work. This paper helps researcher to have a good understanding about various replication techniques and provides broad spectrum of scope for further works in replication in the evolving computing paradigm.

5. References

[1] F. Wang, J. Qiu, et al, "Hadoop high availability through meta data replication", in Proc. First Intl. Workshop cloud data manage, 2009.

[2] A. Gao, l. Diao, "Lazy update propagation for data replication in cloud computing", in Proc. 2010 5th Int. Conf. Pervasive computing and applications, Dec 2010 pp 250-254.

[3] W. Li, Y. Yang, J. Chen et al, "A cost effective mechanism for cloud data reliability management based on proactive replica checking", in proc 2012 12 th IEEE/ACM Int. symp. Cluster, cloud and grid computing, may 2012, pp 564-571.

[4]T. Hara, "Quantifying impact of mobility on data availability in mobile Ad hoc Networks", IEEE Transactions on mobile computing, Vol. 9, No. 2, Feb 2010.

[5] X. jia, D. Li, H. Du and Jinli Cao, "On optimal Replication of data object at hierarchical and transparent web proxies", IEEE Transactions on Parallel and Distributed systems, Vol. 16, No. 8, Aug 2005.

[6] J. H. Choi, K. S. Shim, S. Lee and K. L. WU, "Handling Selfishness in Replica Allocation over a Mobile Ad Hoc Network", IEEE Trans. Mobile Computing, vol. 11, no. 2, pp. 278-291, Feb. 2012.

[7] S. Zaman, D. Grosu, "A distributed algorithm for the Replica Placement Problem", IEEE Transactions on Parallel and Distributed system, vol. 22, no. 9, pp. 1455 - 1468, Sep. 2011

[8] T. Wu and David Starobinski, " A comparative analysis of server selection in content replication networks", IEEE transactions on networking, Vol 16, No. 6 Dec 2008.

[9] Sun DW, Chang GR, Gao S *et al.* "Modelling a dynamic data replication strategy to increase system availability in cloud computing environments. Journal of computer science and technology", 27(2): 256-272 ,Mar. 2012. DOI 10.1007/s11390-012-1221-4

[10] Jenn-Wei Lin, C. H. Chen and J. Morris Chang, "QOS Aware Data Replication for Data Intensive Applications in Cloud Computing Systems", Early Articles, IEEE transaction on

cloud computing, vol. 1, no. 1, pp. 101- 115, 2013, DOI:10. 1109/TCC. 2013. 1 Sep 2013.

[11] Kang Chen,Haiying shen, "Maximizing P2P File access availability in mobile ad hoc networks through replication for efficient file sharing", IEEE Transactions on computers,VOL.64,NO.4,APRIL 2015.

[12] Ahamedin et al, " socila community –partition aware Replica allocation in Ad-hoc social networks", IEEE Intl. Conf. On Green Computing and IEEE Internet of Things and IEEE Cyber, physical and social computing,2013.

[13] Neeraj et al, "Replication aware data dissemination for Vehicular ad hoc networks using location determination",Mobile netw Applications, pp : 251-267,Springer, 2015.

[14] Dejene Boru, " Energy Efficient Data Replication in Cloud Computing Data Centers ",Globecom 2013 workshop - Cloud computing systems, Networks and applications.

[15] Aamir nahir et al, "Replication based load balancing",IEEE Transactions on parallel and Distributed Systems,Vol. 27 ,No.2,Feb 2016.