ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

Distributed File Replication Algorithm for Content based File Sharing System using Dynamic Peer-to-Peer Networks

Kunduru Saidi Reddy[1], Ailuri Venkatrami Reddy[2]
[1,2]*Assistant Professor, Department of IT*
[1,2]*Malla Reddy Engineering College, Hyderabad, Telangana, India*

Abstract

The main aim of the research work is to develop the data replication algorithm based on data security, data processing and load balancing in order to take the minimum energy consumption and high data availability rate in the network. Mobile Ad hoc Networks (MANET) take collection of wireless mobile system dynamically forming new network without the use of any backend infrastructure or centralized administration. Based on Peer-to-Peer (P2P) network file destitute is mainly implementing in MANET. Several data replication protocol is proposed to minimize results degradation. These selfish nodes could data is reduce total data in the network. We introduce a new concept of Distributed File Replication algorithm which considers file dynamics network such as file addition and deletion in dynamic manner. Content-based file sharing scheme using dynamic network is proposed user interest is determined by the proposed scheme before searching and sharing the files in the peer-to-peer network. The locations in the network are utilized as per the contents of the files to be shared. It provides security to decentralized p2p network by the implementation of key server and Intrusion Detection System (IDS) with new P2PHBA algorithm is used for the prediction modify path in the network by the scout is implementation to the efficient file sharing. The present replication protocol drawbacks, they are node storage and the allocation of resources in the replications. Future wireless communications are heading many all-Internet Protocol (all-IP) design and will rely on the Session Initiation Protocol (SIP) to manage services such as voice over IP (VoIP).

Index Terms: *Content-based file sharing; interest extraction, interest-oriented file sharing; peer-to-peer network, Mobile Ad hoc Network (MANET), file Replication, Query Delay.*

1. Introduction

In a mobile ad hoc network (MANET), mobile hosts can communicate directly with one another using direct pair wireless links. Because it requires no fixed infrastructure and most of the time no explicit administration a MANET can extremely useful to support communication in challenging situations, such as in rural, remote disaster-struck areas. P2P computing refers to technology that enables two or more peers to collaborate spontaneously in a network of equals (peers) by using appropriate information and communication systems without the necessity for central coordination.[1] Content inserted into the network is stored and forwarded by cooperating nodes. Metadata and queries are also inserted to represent essential attributes of content and to retrieve appropriate content from the network. Routing

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

and caching perform in-network matching between metadata and queries. Content and metadata/queries must be protected by a decentralized security framework to enable access control of content. Optimization of the content management strategy under constraints can be seen like many other problems in networking as a utility maximization problem. Generally, optimizations at each layer require situation- and resource-aware cross-layer adaptation that is cognizant of features, limitations, and dynamicity at each layer to maintain content accessibility with reasonable trade s between availability and bandwidth. [2].

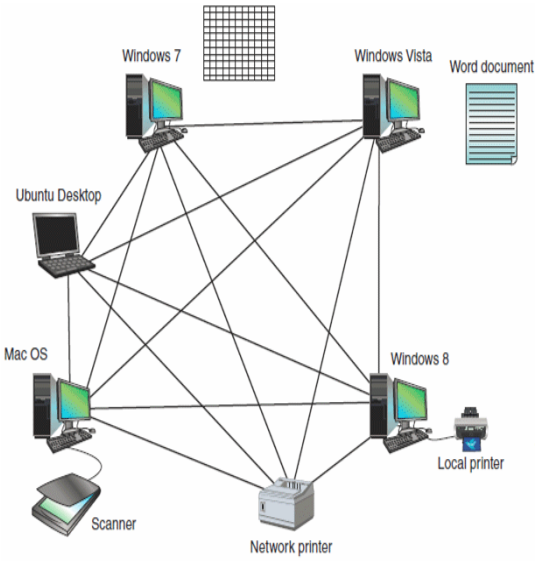

Fig. 1. MANET Peer-to-Peer Networks

For instance, the degree of redundancy for caching of content in a cluster of nodes should take into account the cluster density and stability (lower layer), and at the same time the type and importance of the content (higher layer). MANET (Mobile ad hoc network) is dynamic networks populated by mobile stations. Stations in MANETs are usually laptops or mobile phones. These devices feature Bluetooth or Wi-Fi network interfaces and communicate in a decentralized manner. Mobile ad hoc networks are composed of a set of communicating devices able to spontaneously interconnect without any pre-existing infrastructure for it. Devices in specific range can communicate in a point-to-point fashion. More and more people are interested in mobile ad hoc networks.

Peer-to-peer (P2P) networks and Virtual Private Networks (VPNs) are two typical overlay networks in constructing large scale distributed applications over large networks. Peer-to-peer networks run on top of the Internet. Peer-to-peer networks are distributed systems where the software running on each node provides equivalent functions. A definition of P2P networking is a set of technologies that enable the direct exchange of services or data between computers. Implicit in this definition are the fundamental principles that peers are equals. P2P systems emphasize sharing among these equals. A pure peer-to-peer system runs without any centralized control or hierarchical organization. A hybrid system uses some centralized or hierarchical resources. Peers can represent clients, servers, routers, or even networks [3].

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

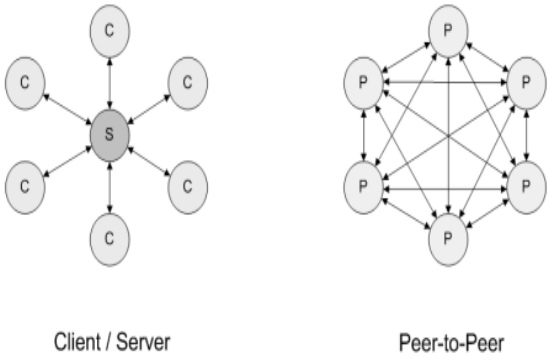

Fig. 2. Client/Server Model versus P2P Model

2. Related Work

With the everyday increasing importance of privacy, security, and wise use of computational resources, the corresponding technologies are increasingly being faced with the problem of file type detection. True identification of computer file types is a difficult task especially when dealing with suspicious goals. Computers are dealing with the huge number of file formats that are transmitting between the insider and outsider networks. Without the true file type detection, the security will not be achievable. File type detection has the most usage and importance in the proper functionality of operating systems, firewalls, intrusion detection systems, anti-viruses, filters, Steganalysis, computer forensics, and applications dealing with the file type classification [4].

Although there are many applications dealing with file type detection, they have very limited methods for determining the true type of files. The newest method of file type detection is based on the file contents. McDaniel and Heydari published the first paper on content-based file type detection [5, 6]. In mobile environment, data is distributed across the network majorly in the form of pushing (data dissemination) and pulling (data hosting) techniques. Few of the possible techniques based on push/pull mechanism have been surveyed in [7]. This includes data dissemination over limited bandwidth channels, location-dependent data querying and advanced interfaces for mobile computers. Moreover, these approaches have few issues to be addressed such as how to store information on database, how to effectively and optimally design database for mobile network, low bandwidth in network; and frequent network partitioning caused by mobility and energy constrained of peers.

The works in [8-10] show the several approaches to address these research issues in mobile networks. Content-based routing and caching solutions for disruption tolerant networking (DTN) require resource provisioning to determine storing or forwarding of a particular piece of content to maximize its availability. Quantifying the benefit and cost of such operations can be formulated as a utility maximization problem Our compositional optimization improves content-based utility by treating individual layers as modules, which makes it easier for further generalization to incorporate various local optimizers, such as deferent routing or caching schemes. Cross-layer optimization under constraints has been studied in networking previously. The work proposes Conquer, a broker-based economic incentive model for

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

mobile peer-to-peer networks. Although the work considers free riders to host data in mobile peer-to-peer networks, it assumes that all peers are trusted and they do not cheat. some strategies for handling selfish behavior have been proposed in the research field of distributed databases. However, these works cannot be directly applied to a MANET, since they did not consider the constraints such as the bandwidth limitation for the detection of selfish nodes and system failures due to frequent node disconnections of a MANET.

3. System Design

Architectural diagram deals with cluster of mobile nodes. In this architecture each cluster contains cluster head. Each cluster head of each cluster is communicated with centralized node. Cluster head is also elected by centralized node. Each member in the cluster is identical in some features. Each member in the cluster requests the cluster head for files and data. Cluster head includes all information about cluster members. Cluster head one group must communicate to cluster head of another cluster in order to get information of cluster member of that particular cluster.

Fig. 3. MANET Architecture

4. Proposed Solution

The content-based file sharing in peer-to-peer networks using threshold. File sharing client allows a number of people to use the same file or files by some combination of being able to read or view it, write to or modify it, copy it, or print it. Typically, a file sharing system has one or more administrators. File sharing clients may all have the same or may have different levels of access privilege. we propose a new distributed file replication protocol to minimize the average querying delay.

Priority Based Dynamic Replication (PBDR) technique is used adding and deleting the replica files based on the priority. The file has been succeeded in the priority competition then adding the replica files, otherwise deletes the replicas. The proposed algorithm well considers the caching overhead and adapts the cache node selection strategy to maximize the caching benefit on different MAC layers. Results show that the asymmetric approach outperforms the symmetric approach in traditional 802.11 based ad hoc networks due to removal of most of the processing overhead. This file sharing client generated request to share file. In request

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

analysis, the request is analyzed to see whether it is authorized or not. If it is authorized then process will be continued otherwise it is stop there. The sample files are collected from different sources to ensure that the sample files of a file type are not generated by one source.

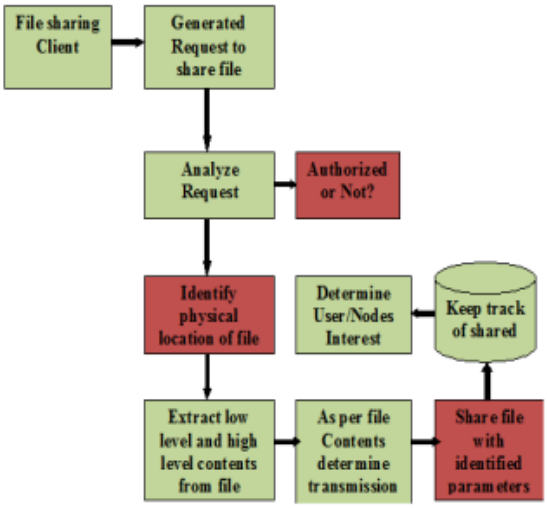

Fig. 4. Content-Based File Sharing using Threshold

A. Replica Allocation

The SCF-tree based replica allocation techniques are inspired by human friendship management in the real world, where each person makes his/her own friends forming a web and manages friendship by himself/herself. He/she does not have to discuss these with others to maintain the friendship. The decision is solely at his/her discretion. After building the SCF-tree, a node allocates replica at every relocation period.

Algorithm:

1. Initialize :
    a. Initialize node memory(in pkt)
    b. Initialize the update timer
2. Send request for allocate replica
3. If node selfish
    a. Give wrong reply or no response
4. If not
    a. Send correct reply
5. If reply is received by originator
    a. Process the reply
    b. And make the route
6. If timer is triggered

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

a. Send the query for checking

7. If node is correspond node then sends the query reply

8. Originator checks the query reply

a. And forms the SCF tree

b. And allocates priority and replica according to the tree

B. The P2PSIP Protocol

P2PSIP implements traditional proxy and registrar SIP functions in a distributed way. Resource information is distributed among all the peers in the overlay, and requests are also handled by the overlay infrastructure. The main advantages of P2PSIP are cost reduction and the elimination of single failure points. Using P2PSIP, any DHT-based P2P network can locate resources in a decentralized way. The clients only use resources from the overlay. They do not participate in its maintenance. The information stored in the P2PSIP overlay consists of registers of the peer nodes and the resources available at these nodes. Kademlia's distributed resource tables (DRTs) are similar to DHTs employed for peer registration A P2PSIP deployment has an overlay name, and the participants can be peers SIP clients. The peers are identified by a peer-ID, and they collectively serve as a directory service for locating resources using the defined DHT structure. The same procedures as those described based on unique resource-IDs and a XOR metric, are employed. In order to keep registers up-to-date. Protocol dSIP [11] uses SIP messages to implement P2PSIP, while preserving the semantics of conventional SIP messages as much as possible. Although there are newer P2PSIP implementations, dSIP has several advantages:

- Simplicity of implementation (still text-based)
- Minimization of the number of protocols required for a P2P UA.
- Easy integration into existing UAs and reuse of available SIP stack implementations.
- Widespread support.

The message to add, remove and query bindings in DHT and resource tables is SIP REGISTER [12]. dSIP supports:

- Peer and resource registration.
- Session establishment.
- DHT maintenance (dSIP is modular, so it allows multiple DHTs).

dSIP peers are active members of the overlay and provide operations to enable self-organization (SIP server-like functions) in addition to the basic functionality of any SIP endpoint. The dSIP overlay serves as a directory service for locating resources.

C. Vehicular Network (VANET)

VANET is one type of mobile ad hoc network where each vehicle act as a router either V2V i.e., Vehicle to Vehicle communication, Vehicle to Infrastructure communication take place. This simply means, Network on wheels. To perform agent migration and mobility

among nodes (handover) in an infrastructure less vehicular ad hoc network (VANET) the concept of opportunistic communication is evaluated and by considering this, [13] looks at location awareness to support agent mobility. One of the interesting applications of this idea is "virtual sensor network" composed of software agents that done the task in form of sensing services with the help of available resources provided by physical nodes like computer platforms, communication devices or physical sensor devices. Exchanging events in vehicular ad hoc network (VANET) helps drivers to receive information about relevant places or to avoid undesirable situations. The work in [14] presents data management as a solution for event exchange in vehicular networks and comparison of two different approaches. From that the first approach is helpful to estimate the relevance of events by calculating of geographic vectors; and second approach exploits digital road maps. For the testing of proposals in a real environment, a prototype is described here. The usefulness of utilizing the information stored in digital road maps for data management and sharing in vehicular networks as well as the first approach is useful in cars where maps are not available has been proved by an exhaustive simulation-based experimental evaluation[15].

D. Dynamic Source Routing Protocol

DSR is a reactive routing protocol with no periodic update messages like table-driven routing protocols to manage a MANET. It is specifically designed for use in multi-hop wireless ad hoc networks where the protocol allows the network to be completely self-configuring self-organizing where there is no need for an existing network infrastructure or administration. Only at the time routing a path is required by a node (On-Demand Routing) the process to find a path is get executed to restrict the excessive use of bandwidth. In DSR the source node (sender, initiator) first determines the communication path to reach the destination node (Source-Routing) and then deposits the intermediate node addresses in the packets. Compared with other reactive routing protocols like SSA or ABR DSR will be beacon less that is there is a no need of hello-messages by the nodes to notify their neighbors about their presence. DSR was developed for MANETs with a moderate speed at small diameter of 5 to 10 hops. DSR uses the Link-State-Algorithms that is each node can be able to save the best way to a destination. Whatever may be changes were made in the network topology will be broadcasted to the whole network by flooding. DSR contains 2 phases Route Discovery (find a path), Route Maintenance (maintain a path).

5. Performance

The output of each technique in the simulation test on NS-2 the hit rates and average delays of the four protocols. We used the following metrics in the experiments:

Hit Rate: It is the number of requests successfully handled by either original files or replica files.

Average delay: This is the average time of all requests that finish execution. The delay that calculate using the throughput and the performance of the requests.

A. Hit Rate

ISSN: 2278-4632
Vol-10 Issue-5 No. 20 May 2020

The hit rates of the four methods with the simulations results. The hit rates continue SAF>DAFN> DCG >PBDR. The PBDR achieve higher hit rate than other methods. since PBDR realizes distributed way, it presents slightly differ from performance compared to others. PBDR considers the intermediate connection properties of disconnected MANETs and replications. DCG only considers temporarily connected group for file replication.

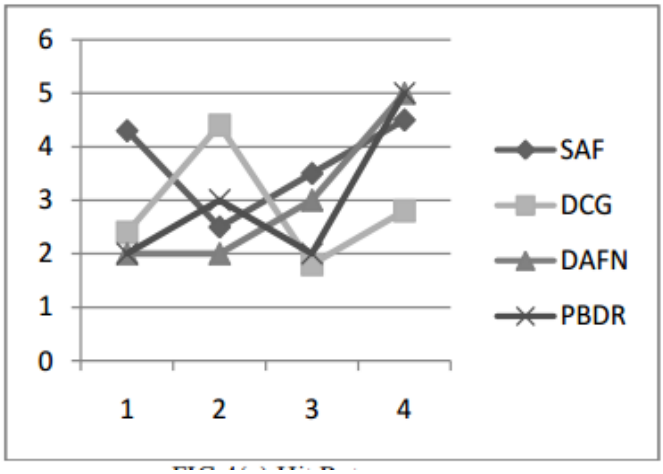

Fig. 5. Hit Rate

B. Average Delay

Demonstrate the average delays of the four methods with simulation results. The average delays show PBDR<SAF<DAFN which is in reverse order of the relationship between the four methods on hit rate as shown in Fig5 . This is because the average delay is related to the overall file availability in descending order. The PBDR have high file availability .SAF distributes every file to different Nodes while DCG only shares data among simultaneously identify neighbor nodes, and DAFN has a low file availability since all files receive equal amount of memory resources for replicas. The PBDR has the minimum average delay in the simulation results.

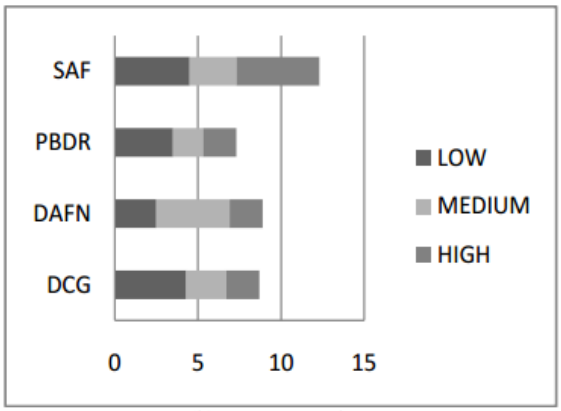

Fig. 6. Average Delay

6. Conclusion

We also plan to improve the current leveling technique by considering the frequency of disconnections and/or the weighted ratio of total number of items and average of shared memory space we will discuss the important research issues in more detail related to mobile databases. This survey presents various mobile applications, including facility of data sharing and data storing for easy and handy access to the static and dynamic information in the mobile databases. It also discussed the various communication strategies in mobile ad hoc networks. The proposed mechanism can be used in any network environment having limited number of resources and maximum number of transmissions. This cluster form of monitoring will also decrease the overhead which may be happened while cluster head has maintaining the routes. IDS will identify the nodes as trusted and untrusted based on some threshold values and also inform about entrusted to all nodes in the network to transfer data in a secured manner. For the dynamic routing purpose nodes outside the network coverage need to get authenticated by IDS. We are currently working on the impact of data updates and different moving patterns on our scheme. In future the proposed scheme can be extended for identifying the requirements of the network for best effort network services

7. Future Work

Future work consists in enhancing protocol with quality of service mechanisms and providing measurement studies of the protocol performance. Considering the problem of overlay topology design, future work could also focus on studying the overlay topology creation and adaptation in case of unknown traffic demands. A hybrid network with a mix of selfish and cooperative nodes is an additional interesting scenario. Heterogeneous values of the overlay cost coefficient could be proposed for each node in the network, and its effect on the overlay topology creation could be studied. Otherwise, there can be an interesting joint between overlay networks and Pocket Switched Networks which could be studied in depth to make solid proposals.

References

[1] “Qik,” http://qik.com/, 2014.

[2] A. Balasubramanian, B. Levine, and A. Venkataramani. DTN routing as a resource allocation problem. In Proc. Conf. on Applications, Technologies Architectures, and Protocols for Computer Communications, SIGCOMM ’07. ACM, 2007

[3] M. Hofmann and L.R. Beaumont, (2005) “Content networking: architecture, protocols, and practice”. Morgan Kaufmann, ISBN: 1558608346.

[4]. Dunham J. G., Sun M. T., Tseng J. C. R. Classifying File Type of Stream Ciphers in Depth Using Neural Networks. The 3rd ACS/IEEE International Conference on Computer Systems and Applications, 2005

[5]. Karresand M., Shahmehri N. File Type Identification of Data Fragments by Their Binary Structure. Proceedings of the IEEE Workshop on Information Assurance, 2006, p.140-147

[6]. Zhang L., White G. B. An Approach to Detect Executable Content for Anomaly Based Network Intrusion Detection. IEEE Parallel and Distributed Processing Symposium, 2007, p.1-8

[7] D. Barbará, "Mobile Computing and Databases-A Survey," IEEE TKDE, vol. 11, no. 1, pp. 108–117, 1999.

[8] D. Chan and J. F. Roddick, "Context-sensitive mobile database summarisation," Proc. ACSC, pp. 139–149, 2003.

[9] N. Tolia, M. Satyanarayanan, and A. Wolbach, "Improving mobile database access over wide-area networks without degrading consistency." ACM, 2007, pp. 71–84.

[10] K. A. Qureshi, S. Mohiuddin, A.-U. Aziz-Uddin, and A.-U.-R. Atique- Ur-Rehman, "A strip-down database for modern information systems," Proc. Computers, pp. 81–88, 2010.

[11]. Cirani S., Veltri L. A Kademlia-Based DHT for Resource Lookup in P2PSIP.

[12]. Wauthy J.-F., Schumacher L. Implementation and Performance Evaluation of A P2PSIP Distributed Proxy/Registrar. Proceedings of the 2007 International Conference on Next Generation Mobile Applications, Services and Technologies

[13] K. Ahmadian, M. Gavrilova, and D. Taniar, "Multi-criteria optimization in GIS: continuous k-nearest neighbor search in mobile navigation," Proc. of Computational Science and Its Apps, pp. 574– 589, 2010.

[14] E. P. De Freitas, T. Heimfarth, F. R. Wagner, A. M. Ferreira, C. E. Pereira, and T. Larsson, "Geo-aware handover of mission agents using opportunistic communication in VANET," Proc. ruSMART/NEW2AN, pp. 365–376, 2010.

[15] T. Delot, S. Ilarri, N. Cenerario, and T. Hien, "Event sharing in vehicular networks using geographic vectors and maps," Mobile Information Systems, vol. 7, no. 1, pp. 21–44, 2011.

