

Content Caching and Replication Schemes for Peer-to-Peer File Sharing in Wireless Mesh Networks

S. Vijaya Lakshmi , M. Sreevani,

CSE Department, Mahatma Gandhi Institute Of Technology

Hyderabad, India

svijayalakshmi1111@gmail.com

Abstract— Multiple nodes coordinate with one another to cache data and share the cached data. Cached data not only replies local access, but also replies data request issued from other nodes. Wireless Mesh Networks (WMNs) have emerged as an important technology in building next generation fixed wireless broadband networks that provide low cost Internet access for fixed and mobile users. Reduce the number of hops that request/data need to travel in the network. An orthogonal evolution in computer networking has been the rise of Peer-to-Peer (P2P) applications such as P2P file sharing. It is of interest to enable effective P2P file sharing in this type of networks. . Finally, we compare the performance of our proposed schemes against other existing schemes using simulations. We focus on P2P file sharing but other applications can also be supported by the proposed schemes. Our main contribution in this paper is innovative schemes for content caching and replication at mesh routers that enhance the performance of P2P file sharing in WMNs. We first motivate our proposed schemes by showing the impact of caching P2P content at mesh routers on the performance of P2P file sharing in WMNs. We then describe the design and operation of our content caching and replication schemes.

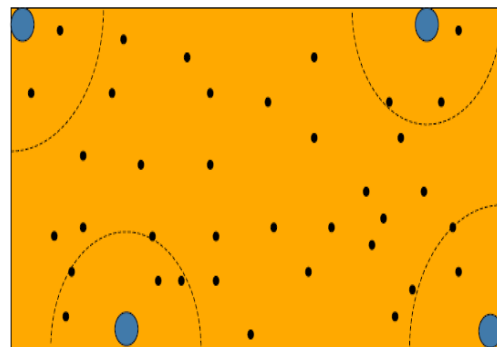
Keywords: *Wireless mesh networks, peer, cache, gateway, router, protocol.*

INTRODUCTION

Info stations are deployed to provide info, such as maps, attractive sites, or restaurant info to mobile users. Users may relay for each other to serve those not directly covered by the info stations. Wireless Mesh Networks (WMNs) are composed of wireless stationary access points (mesh routers) that facilitate the connectivity and intercommunication of wireless mobile clients (mesh clients) through multi-hop wireless paths. Users also share data got from the infestations. The dominant design approach for WMNs is a two-tier architecture, wherein an access tier connects mobile end-user computing terminals (mesh clients) to stationary mesh nodes (mesh routers), and the mesh routers form a multi-hop wireless backhaul tier that routes data packets between mesh clients within the WMN and between mesh clients and few number of gateway nodes that are wired to the Internet [1] (Fig. 1).

Many major operators have already considered the wireless mesh network for their wireless city initiatives (Birmingham and Newcastle wireless city initiatives [2]). A promising example of WMNs deployment is the wireless community networks (Houston [3]). Users in a wireless community

network (neighborhood, campus, etc.) share their communication facilities (wireless access points) and form a wireless mesh network to be used by community members. Wireless community networks can be used to provide innovative services for end users in a community (e.g. Voice over Internet Protocol (VoIP), streaming, and Peer-to-Peer (P2P) services [1]). Similar to the Internet and as the mentioned innovative services suggest, we project forward and expect P2P file sharing traffic to consume a significant amount of WMNs capacity. Users also share data got from the info stations.



In prior work, we designed a network topology aware system (P2PMesh) that enables efficient P2P file sharing in WMNs. The P2PMesh system benefits from the support for the P2P applications at mesh routers and achieves significant performance improvement [4]. We now compliment our prior work and propose schemes for content caching and replication at mesh routers that further enhance the P2PMesh system performance. Cooperative caching in wireless network has been studied at a very high level and the evaluation is limited by pure simulation. Implementation of cooperative caching has only be seen in Internet and Web environment. The particular characteristics of WMNs such as the existence of stationary mesh routers, which typically have a very low probability of leaving the network and have high upload capacity compared to end users, motivated us to propose schemes for content caching and replication at mesh routers. The proposed content caching scheme benefits from the protocol proposed for data dissemination in the P2PMesh system [4]; while the proposed content replication scheme exploits the P2PMesh system architecture and takes into consideration the variation of files popularity and files replicas in the WMN over time.

The problem of minimizing the access cost of a file in P2P file sharing is more serious in WMNs compared to wired networks. The file access cost, in the context of multi-hop networks, is defined as number of wireless hops between a peer who is downloading a file and the nearest file replica [5]. Due to contention for the wireless media between neighboring mesh routers and interference between traffic on adjacent wireless links, file downloading rate at peers is significantly reduced when data traverses a long path in a WMN [6]. Minimizing the file access cost implies less bandwidth and power consumption in WMNs. Content caching is a common technique used to improve the performance of data access by increasing number of file replicas in the network.

There is limited storage capacity at mesh routers and hence a mesh router on the route between a file provider and a downloading peer cannot cache all files which it forwards to next hop node towards the destination. The proposed content replication scheme computes the exact number of replicas needed for each file available in the WMN such that the average access cost for all file in the WMN is minimized and the storage disks at mesh routers are fully utilized. The scheme enables mesh routers on the route between a file provider and a downloading peer to collect sufficient information about the transmitted file that allow them to take local decision whether to cache the transmitted data without introducing large overhead to the WMN.

We motivate our work by showing the performance enhancement when caching content at mesh routers in terms of average data rate at downloading peers (Section III). We describe the design of our content caching and replication schemes in Sections III and IV respectively. We compare the performance of the proposed schemes against other existing content replication schemes in Section V.

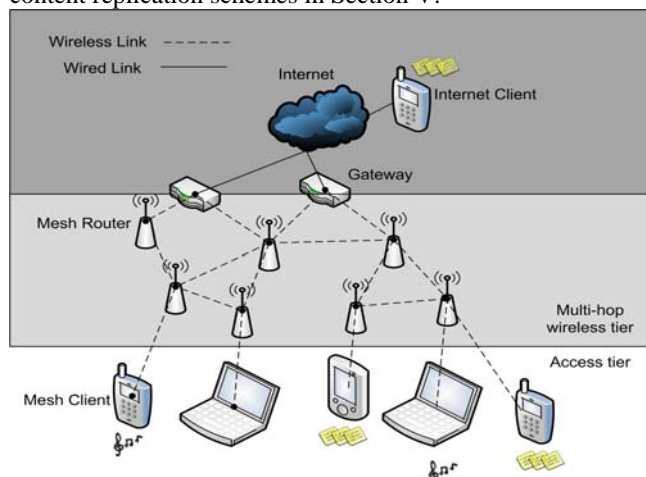


Fig. 1. Wireless Mesh Network Architecture

EXISTING SYSTEM

Most caching protocols in multi-hop wireless networks have been proposed for Ad hoc networks [7], [8]. Cao et al. proposed two schemes for caching at Ad hoc networks: Cache Data which caches the data, and Cache Path which

caches the data path at nodes on the route between source and destination [7]. However, the characteristic of WMNs call for different caching designs. A few protocols have been proposed for content caching and replication in WMNs [9], but these schemes do not exploit the broadcast nature of wireless links in WMNs and only enable mesh routers that are located on the route between source and receiver to cache data. However, the broadcast nature of wireless links allows mesh routers in the route vicinity in addition to mesh routers on the route to the destination to overhear the wireless transmission and cache data.

Many content replication algorithms have been proposed to improve the performance of P2P file sharing. Performance metrics such as file query cost (number of peers that a file requester should inquire before the required key is resolved) and file access cost (distance between the location of the closest required content replica and the downloading peer in terms of number of physical hops) were considered. Cohen et al. proposed an optimum replication method (square-root rule) to minimize the file query metric in unstructured P2P file sharing networks, which replicates a content at end users such that the density of a content in the overlay network is proportional to the square root of the content popularity [10]. Jin et al. propose an optimum content replication strategy specialized for WMNs (minimum access strategy), that minimizes the file access metric [11]. The strategy considers a set of files in a large WMN where each file has fixed probability of being requested (popularity) and computes the optimal number of replicas needed for each file in the set. Jin et al. have shown that the optimal strategy replicates a file i such that number

of file i replicas is proportional to $p_i^{0.667}$, where p_i is file i popularity in the WMN. The drawbacks of the minimum access strategy are the following: (i) it considers a finite set of files in the WMN with each file having fixed popularity at all time. However, files popularity and the number of files available in the the WMN varies over time [5]. Therefore, an adaptive content replication strategy is needed. (ii) it requires knowledge about files popularity and the number of replicas for each file in the network. Jin et al. did not propose any mechanism to address these problems. Our proposed content replication scheme addresses these problems and benefits from the structure of P2PMesh system. The P2PMesh system decouples mesh clients from participating in the P2P overlay network. Only mesh routers which participate in the P2P file sharing application are involved in the overlay network. Therefore, each mesh routers in the P2PMesh system monitors file requests rate (file popularity) and the number of file replicas in the WMN for each file's descriptor it is responsible for [4].

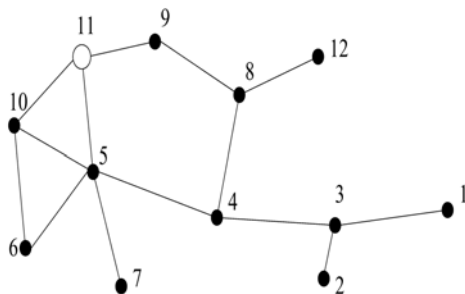
PROPOSED SYSTEM DESIGN OF THE CONTENT CACHING SCHEME

A WMN is typically spans a small geographic area and mesh routers, therefore, can be easily over-provisioned with CPU, memory, and disk storage. Therefore, it is reasonable to cache

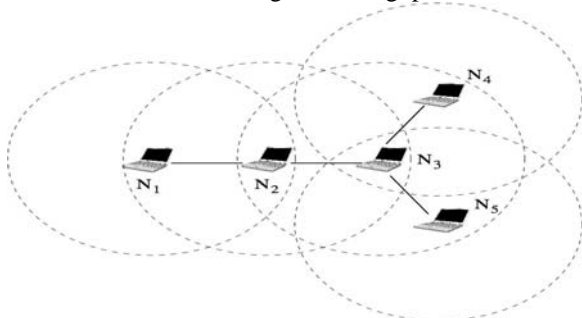
P2P files at mesh routers and upload cached files from mesh routers directly to downloading peers.

To demonstrate the impact of caching content at mesh routers on the performance of P2P file sharing in WMNs, we simulated a network consisting of 98 mesh clients (peers) with 1Mbps upload bandwidth and 49 mesh routers with 54Mbps upload bandwidth. 980 distinct files were distributed on the end users (peers) uniformly at random such that each peer caches 10 distinct files. All files are with same size (1Mbyte) and the packet size was fixed at 1024 bytes. Distance between neighboring mesh routers was 100m and coverage range for each mesh router was 250m. Background traffic was coexisted in the network with P2P traffic. To simulate background traffic, we selected 10 pairs of source=destination uniformly at random and established constant bit rate (CBR) flows between them with rate 100Kbps. Each peer requests a file at rate exponentially distributed with mean 10 sec. A downloading peer selects two file providers for each requested file. The performance metrics for evaluation were: average received data rate at downloading peers and the average number of wireless transmissions (packets) needed to transmit a file from selected file providers to the destination. We considered the following two cases:

Case 1: Peer-to-Peer, where each peer caches 10 distinct files and number of replicas. Downloading peers select file providers with the shortest paths.



Case 2: P2P with mesh routers support, where files were distributed on peers such that each peer caches 10 distinct files and number of file replicas (as in case 1). Number of mesh routers (K) were selected uniformly at random to assist peers in uploading data. Each selected mesh routers caches files it overhears during 1 hour simulation time and upload data it caches to other peers in the network upon receiving a file request. A downloading peer selects two file providers from which it receives the highest throughput.

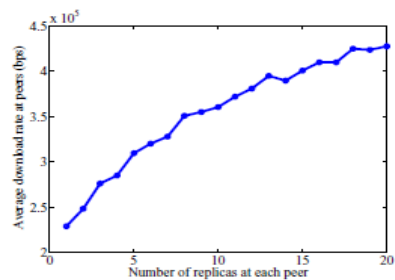


WIRELESS MESH NETWORK

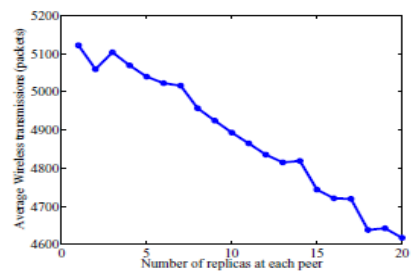
In case 1, we observe increasing trend between number of replicas stored at peers and average downloading rate at peers (Fig. 2(a)). This is because when number of replicas at peers is high, the probability that a downloading peer locates the required file at a neighboring peers is high. This result in higher received data rate at downloading peers and less number of wireless transmissions as file’s segments traverse shorter paths from file providers to destinations (Fig. 2(b)).

In case 2, we observed that no matter how large is the number of replicas stored at end users (peers), downloading peers always download required files from the participating mesh routers. Despite the fact that neighboring peers, in terms of number of wireless hops, can serve as file providers for the downloading peers, the received throughput is higher when a downloading peer selects mesh routers as file providers (Fig. 3(a)).

We conclude that the performance of P2P file sharing in WMNs, in terms of file downloading time, is constrained by the uploading bandwidth at file providers and, therefore, no matter how good the files are replicated at end users (peers), the received data rate at downloading peers is limited.



(a) Average received data rate for varied number of replicas stored at peers

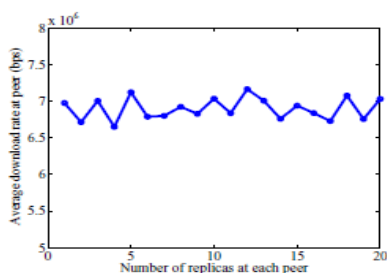


(b) Average number of wireless transmissions for varied number of replicas stored at peers

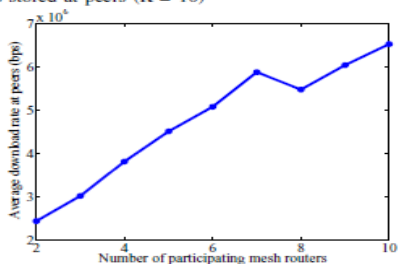
Fig. 2. Peer-to-Peer case

Another interesting observation is that increasing number of participating mesh routers has a significant impact on the data rate received at downloading peers (Fig. 3(b)). Although mesh routers only cache files that they overhear, the average downloading data rate at clients increases more than five times compared to case 1 when only two mesh routers participate in the P2P file sharing (K=2). These results suggest strategies for content caching and replication at mesh routers, which have much higher upload bandwidth compared to end users.

We now describe the design of our proposed caching protocol. Due to the difficulties in modifying the routing module to integrate the caching function into the network layer [8], we propose a layered design that decouples the caching module from the routing module. Specifically, we propose to add a sub-layer to the set of layers defined by the IP protocol suite directly above the network layer. The proposed caching protocol works as follows: suppose mesh router Y is the file provider for the downloading peer P (Fig. 4). Suppose the network routing protocol (e.g. DSR [12]) computes the route between Y and P as Y-B-C-X-P. File provider Y then uses the IP layer to broadcast the file's packets required by P after appending in the sub-layer the complete route information to P and information about the transmitted file such as file's popularity and the number of optimal replicas required for the file in the network (Fig. 5). When mesh router B overhears the broadcasted packet, it scans (sniff) the sub-layer and when it finds its IP address in the sub-layer's route information field, it rebroadcasts the packet. Next hop mesh router on the route specified in the sub-layer (node C) does the same until the packet reaches the destination P.



(a) Average received data rate for varied number of replicas stored at peers (K = 10)



(b) Average received data rate for varied number of participating mesh routers (number of replicas stored at peers = 100)

Fig. 3. P2P with mesh routers support case

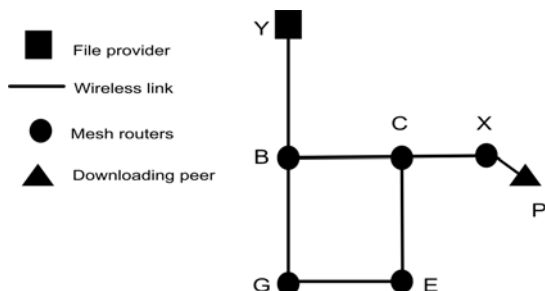


Fig. 4. An Example of P2P file sharing setting.

Only mesh routers which are on the route to the destination rebroadcast the packet, while all mesh routers which overhear the data transmission such as B, C, G, E and X may cache data. We distinguish between two kinds of caching at mesh routers which overhear the file transmission: network-demand caching, where a mesh router that connects the downloading peer to the network (i.e. node X in Fig. 4) receives an explicit request to cache the file; and optional caching, where a mesh router that overhears the transmission decides to cache data. If all mesh routers on the route to the destination decide to cache data, many replicas of the transmitted file will be concentrated in a small physical deployment area. To avoid this problem, we add cache-counter field to the sub-layer. The cache-counter is increased by one every time an intermediate mesh router on the route to the destination optionally decides to cache the transmitted file. A mesh router on the route does not cache data unless cache-counter=0.

Application			
TCP			
Total sub-layer length	Key	File segments IDs span	
Cach_Count	File size	File popularity	Required file replicas
Route to 1 st peer	Route to 2 nd peer	Route to N th peer
IP Layer			
Data Layer			

Fig. 5. Sub-layer Format

DESIGN OF THE CONTENT REPLICATION SCHEME

Each participating mesh routers in the P2PMesh system monitors the file popularity and the number of replicas in the WMN for each file's descriptor it is responsible for [4]. Specifically, each mesh router continuously counts the number of file requests it receives for each file's descriptor within a specific period of time (T).

Let M be the total number of participating mesh routers in the P2P file sharing, S be the maximum number of files that can be stored at a mesh router's storage disk (assuming mesh routers have the same storage capacity and assuming that all files are of the same size), A be the average total number of requests for all files in the WMN within time period T, t_i be the number of requests for the i-th file in time period T, and N be the total number of distinct files in the network. We assume that $M \cdot S \gg N$.

The popularity of the i-th file at time t is $p_i(t)$, where $p_i(t) = t_i / A$. The optimal normalized density of replicas required for the i-th file in the WMN at time t is $d_i(t)$, where $d_i(t) / p_i(t) = 0.667$ [11].

Each mesh router computes p_i for each file it is responsible for every T time period and sends this information to a predefined node in the network (content replication manager).

The content replication manager upon receiving this message from all participating mesh routers, computes P_i and returns this value to all participating mesh routers. Each mesh router then computes R_i and D_i for each file it is responsible for as:

$$R_i(t) = \frac{M \times S \times p_i(t)^{0.667}}{\sum_i p_i(t)^{0.667}}; \tag{1}$$

where $R_i(t)$ is the number of optimal mesh routers in the WMN that need to cache the file i at time t such that ($R_i \leq M$); and

$$D_i(t) = R_i(t) - E_i(t); \tag{2}$$

where $E_i(t)$ is the number of mesh routers that have file i available in their storage disk at time t ; and $D_i(t)$ is number of extra mesh routers that need to cache the file i at time t to achieve the optimum replication in terms of file access cost. Every time a mesh router receives file replacement message for a file i it is responsible for, it decreases E_i value by one. When the downloading peer selects a file provider and establishes a route to the selected file provider, the file provider uses the data transfer scheme described in Section III and transmits the file's packets after appending in the sub-layer the following information: cache-counter = 0, file popularity (p_i), and D_i .

Algorithm 1 Pseudo code for the file replacement algorithm.

```

{At downloading peer's mesh router;}
if  $D_i = 0$  then
    {optional caching}
    if there is free space in mesh router storage disk then
        cache the file;
    else
        find a file optionally cached with least access cost and if the incoming
        file has higher access cost, replace the file by the incoming file;
    end if
end if
if  $D_i > 0$  then
    {network-demand caching}
    if there is free space in mesh router disk then
        Cache the file;
    end if
    if there is no space for incoming file  $i$  and cache-counter=0 then
        if there are optionally cached files in mesh router disk then
            Evict the file optionally stored at mesh router disk which has least
            access cost and replace it with new incoming file;
        end if
        if all files cached at mesh router are network-demand then
            Find a stored file which least uploaded from the mesh router to the
            peers within time  $T$ ;
            send file-evict message to the mesh router which is responsible
            for the evicted file;
            if the file eviction request is approved then
                Evict the file and replace it with new incoming file;
            else
                find the second least requested file and repeat the process;
            end if
        end if
    end if
    if there is no space for incoming file  $i$ , all files cached at mesh router
    are network-demand, and cache-counter>0 then
        Contact the mesh router node which cached the file and inform it
        about the network-demand caching;
    end if
end if
end if

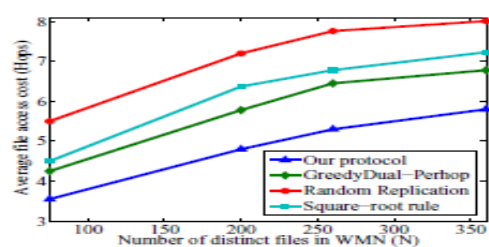
```

When the packet reaches the downloading peer's mesh router (node X in Fig. 4), node X scans D_i field in the sub layer (Fig. 5). If $D_i > 0$, node X caches the file (network demand caching). If X's storage disk is full, it checks if any mesh

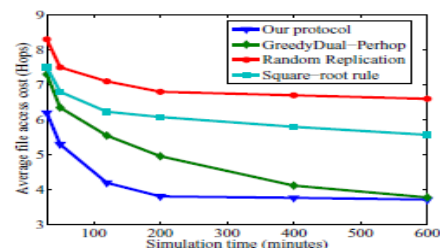
router on the route cached the transmitted packet by scanning the cache-counter field in the packet's sub-layer. If cache-counter>0, this implies that a mesh router on the route optionally cached the transmitted packet and it is not necessary for node X to cache data. However, node X has to send a message to the mesh router which cached the file to confirm that the cached file is now considered as network-demand replica. On the other hand, if cache-counter=0, node X uses the file replacement algorithm (Algorithm 1) as follows: node X evicts the file with the least file access cost which was optionally cached. If all cached files at the X's storage disk were network-demand replicas, mesh router X evicts the file j which least uploaded from the mesh router X to peers in the WMN within time period T (Least Recent Used). Before it evicts file j and replaces it with the incoming file, it sends a file-evict message notification to the mesh router which is responsible for the file j . Upon receiving this message at the mesh router which is responsible for the file j , it computes the number of mesh routers that need to store the file j in the WMN at time t (i.e. $R_j(t)$) and if this number of is maintained, it confirms the eviction of file j .

PERFORMANCE EVALUATION

We simulated, using Opnet, a mesh network consisting of 200 stationary mesh routers. Each mesh router was connected to four other mesh routers on average. Mesh routers were equally separated with distance 100m. Each mesh router was equipped with 802.11b radio with link rate 11Mbps and coverage range of 250m. DSR [12], one of the most popular network routing protocol for wireless multi-hop networks, was employed. N distinct and equal size files were distributed at mesh routers in the WMN uniformly at random. Rate of requesting each file (file popularity) was different. Each mesh router can store only 3 files in its storage disk and each mesh router were made to request a file at intervals exponentially distributed with mean 10 minutes.



(a) Average file access cost for varied N (1 hour simulation time).



(b) Average file access over simulation time ($N=260$).

Fig 5

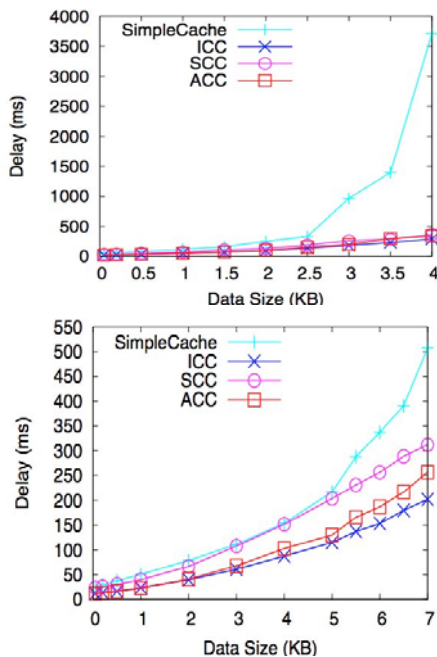


Fig. 6. Content replication schemes performance comparisons

We carried out simulations each for one hour to compute average file access cost for varied number of distinct files in the network (N) using the following content replication protocols: 1- Our content replication scheme. 2- Random content replication, where the downloading peer's mesh router randomly decides to cache the file segments and if its storage disk is full, it randomly replaces a file stored in its disk with the incoming file (Random Replication). 3- Adaptive content replication (Greedy Dual-Per-hop) [9], where downloading peer's mesh router upon receiving the packet, compares the file access cost of the incoming file with those stored in its storage disk and evicts the file with least file access cost from its store and replaces it with the incoming file. 4- Square-root content replication strategy [10], which replicates the i -th file in the network such that d_i is proportional to p_i (Square-root rule).

CONCLUSION

We observe that for all investigated content replication protocols, when number of distinct files in the network increases, files access cost increases (Fig. 6(a)). This is because when more distinct files are available in a network with limited nodes storage capability, replicas for each file decreases. However, our protocol outperforms other content replication protocols in terms of average files access cost because it optimally replicates files in the WMN, avoids bias against less popular files, adapts to varying files popularity in the WMN, and utilizes the available storage at mesh routers. We also computed the average files access cost for varied simulation time in all content replication protocols under investigation when $N=260$. An interesting observation is that when simulation time is short, the difference in file access cost between Greedy Dual-Per-hop and our protocol is

relatively high. However, the difference decreases with time and the performance of both protocols converges when mesh routers' storage disks become fully utilized (Fig. 6(b)). The reason behind this behavior is that Greedy Dual protocol approximates the optimal content replication strategy only when the network is stable (i.e. the popularity of files does not change over time, and no new files are introduced). In the contrast, our protocol, which enables mesh routers to obtain fresh information about the file replicas and popularity of files, enables better adaptive content replication in a dynamic P2P file sharing over WMN. Moreover, our protocol enables all mesh routers in route vicinity which can overhear the relayed file frames to optionally cache data and, therefore, allows for faster storage disks utilization at mesh routers.

REFERENCES

- Rüdiger Schollmeier, A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications, Proceedings of the First International Conference on Peer-to-Peer Computing, IEEE (2002).
- Lua, Eng Keong; Crowcroft, Jon; Pias, Marcelo; Sharma, Ravi; Lim, Steven (2005). "A survey and comparison of peer-to-peer overlay network schemes"
- OECD Organization for Economic Co-operation and Development (July 24, 2008). OECD Economic Surveys: South Africa 2008: South Africa - Economic Assessment. OECD Publishing. p. 103. ISBN 978-92-64-04692-4.
- J. Jun, M.L. Sichitiu, "The nominal capacity of wireless mesh networks", in IEEE Wireless Communications, vol 10, 5 pp 8-14. October 2003.
- S.M. Chen, P. Lin, D-W Huang, S-R Yang, "A study on distributed/centralized scheduling for wireless mesh network" in Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing, pp 599 - 604. Vancouver, British Columbia, Canada. 2006.
- Pathak, P. H.; Dutta, R. (2011). "A Survey of Network Design Problems and Joint Design Approaches in Wireless Mesh Networks". IEEE Communications Surveys & Tutorials 13 (3): 396-428. DOI:10.1109/SURV.2011.060710.00062.
- V. Kawadia, P. R. Kumar (February 2005). A Cautionary Perspective on Cross-Layer Design in IEEE Wireless Communications. pp. 3-11.
- "FRS and Sysvol Improvements". What's New in Group Policy in Windows Vista and Windows Server 2008. Microsoft TechNet. Retrieved 2006-08-04.[dead link]
- A Windows Vista-based client computer does not participate in the replication process if you add the client computer to a DFS Replication service group by using the DFS Management snap-in on a Windows Servers 2003 R2-based computer
- <http://blogs.technet.com/b/filecab/archive/2008/02/08/sysvol-migration-series-part-1-introduction-to-the-sysvol-migration-process.aspx>
- Verifying File Replication during the Windows Server 2008 DFSR SYSVOL Migration -Down and Dirty Style, Ask the Directory Services Team, Microsoft TechNet. Accessed
- David F. Bantz and Fred'eric J. Bauchot. Wireless LAN Design Alternatives. IEEE Network, 8(2):43-53, March/April 1994.
- Vaduvur Bharghavan, Alan Demers, Scott Shenker, and Lixia Zhang. MACAW: A Media Access Protocol for Wireless LAN's. In Proceedings of the ACM SIGCOMM '94 Conference, pages 212-225. ACM, August 1994.
- Robert T. Braden, editor. Requirements for Internet Hosts - Communication Layers. RFC 1122, October 1989.
- Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. In Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'98), pages 85-97, Dallas, TX, October 1998. ACM.