

An effective Caching on Forwarding Table Scheme for Metro Ethernet

Xiaocui Sun and Zhijun Wang

Department of Computing, The Hong Kong Polytechnic University, Hong Kong

Email: csxcsun@comp.polyu.edu.hk

Abstract

Metro Ethernet has recently attracted remarkable attention due to its simplicity and low cost. Although Ethernet is the dominant technology in the Local Area Network (LAN) which usually handles small number of nodes, when deploying in Metropolitan Area Networks (MANs), it faces certain challenges. One major challenge is its broadcasting based address resolution scheme. A broadcast-based address resolution scheme triggers many broadcast messages in a network, and hence make the Ethernet unscalable. This paper proposes a cache scheme, called Caching on Forwarding Table (CFT), to reduce broadcast messages for Metro Ethernet. CFT caches end user's Address Resolution Protocol (ARP) entries by inserting an IP address in each forwarding entry. If an ARP request is answered by a cached entry, the broadcast is stopped. Hence CFT can significantly reduce the broadcast messages in the network. We also discuss how to cache ARP entries in a Metro Ethernet using End User Enabled MAC-in-MAC (MiM) scheme. The simulation results show that CFT can effectively decrease the communication messages for address resolution and make Metro Ethernet more scalable.

Keywords: *Metro Ethernet Cache Forwarding Table Scalability*

1. Introduction

Ethernet is a dominant layer 2 communication technology. Currently, it has gained popularity for deploying in Metropolitan Area Networks (MANs). The traditional MANs are generally built based on layer 3 IP technology, usually using Asynchronous Transfer Mode (ATM) as its layer 2 transport, whereas the Local Area Networks (LANs) are dominated by layer 2 technologies, mostly based on the Ethernet technology. As a result, when a data frame is sent across a MAN, it may have to travel through multiple heterogeneous networks, and hence it may need to be re-encapsulated in different frame formats a number of times and even be split into multiple smaller frames/cells or merged with other frames/cells into a larger frame. This not only makes the packet forwarding complex and inefficient, but also adds complexity and cost to the router/switch design, as well as network operation and management, in support of heterogeneous networks and interfaces. Clearly, deploying the same technology, such as Ethernet, in both MAN and LAN segments can potentially reduce the complexity and cost in network design and management, and improve packet forwarding performance.

Metro Ethernet has been highly attractive due to its cost effectiveness and easy management. However, it has to solve the scalability issues [3] [4] [6]. Ethernet was originally designed as a Local Area Network (LAN) technology that usually handles small number of users. Broadcast is frequently used to resolve unknown MAC addresses. Some protocols such as ARP [7] and DHCP [8] also use the broadcast service as a service discovery mechanism. For example, to solve an MAC address for an IP address, an ARP request message is broadcast through the network so that the corresponding end user can receive and

response an ARP reply message. The broadcast based address resolution schemes make Ethernet extremely convenient and easily accomplished. However, for MANs with millions of end users [9], high frequency broadcast messages waste a lot of bandwidth for address resolution. Ethernet switches have the ability to automatically learn the location of end users by checking the frames they received and record the information on its forwarding table. Frequently broadcast frames accelerate the replacement of the table entries which may leads to forwarding table explosion and hence triggers excessive frame flooding. Moreover, an end user has to take resource to handle every broadcast message [10]. Virtual LAN (VLANs) [22] technology is one way to reduce the broadcast message by logically segmenting the whole network into separate communication groups. However, even after segmentation, the number of users located in one VLAN is still huge in a MAN, making the Metro Ethernet unscalable.

In this paper, we propose a cache scheme, called Caching on Forwarding Table (CFT), to reduce broadcast messages for Metro Ethernet. In CFT, an IP address is inserted into its corresponding forwarding entry, and hence a forwarding entry is a cached entry of an Address Resolution Protocol (ARP) entry. When a node receives an ARP request message, the cached entry can be used to answer the request and then the broadcast is stopped. Hence CFT can significantly reduce broadcast messages in a network. CFT does not change the forwarding table maintenance. When a node receives an ARP request or response message, the source MAC address and source IP address are recorded in the forwarding table. CFT is implemented in the forwarding nodes, particularly in Customer Edge (CE) nodes and Provider Edge (PE) nodes, but it does not have any impact to end users. We also discuss how to use CFT in a Metro Ethernet using End user enabled Mac-in-Mac (EMiM) [34] encapsulation Scheme. In EMiM, the forwarding table in a PE node does not record end user's forwarding entry, and a cache is allocated in a PE node to reduce the broadcast messages. An analytical model is conducted to evaluate the hit ratio of CFT. The experiment results show that CFT can reduce up to 50% broadcast messages.

The rest of the paper is organized as follows: Section II gives the related works. Section III presents the details of CFT and how to do cache in the EMiM scheme. An analytical model is conducted in section IV. The performances are illustrated in Section V. Finally, Section VI concludes the paper.

2. Related work

Scalability is a major challenge for deploying the Ethernet technology in MANs. There are a lot of works focused on the development of scalable solutions for Ethernet. VLAN technology [21] is a useful scheme to make Metro Ethernet scalable. It uses a VLAN tag [22][4] to partition a single large Ethernet to multiple VLANs. The broadcasting messages for address resolution are limited to a single VLAN instead of the whole network, and hence the redundancy traffic is reduced. The VLAN tag has only 12 bits and hence can only support up to 4096 active VLANs at any time in a network. To support more active VLANs, Q-in-Q or VLAN stacking encapsulation [23] is proposed. In this scheme, a PE node inserts an additional Q-tag in the Ethernet frame to support more active VLANs. However, this scheme still cannot avoid learning table explosion. MAC-in-MAC (MiM) encapsulation [4][13] can be used to reduce the forwarding table size in the CNs, but the PE nodes still need to keep a potential large number of forwarding entries for MiM encapsulation.

Some hash based address resolution schemes [24] [25] [26] are proposed to eliminate the reliance broadcasting frame learning. Instead, when a destination MAC address is missed in the forwarding table, the frame is routed to a designated user based on the MAC hash value. But routing frame with unknown destination addresses to designated user may take the frame travel more unnecessary hops to the destination, and make the traffic control more difficult.

SmartBridge [12] allows finding the shortest forwarding path by exchanging topology information among bridges. Hence a full knowledge of the network topology should be obtained. In [27], a MAC address translation scheme is proposed for frame forwarding. The flat MAC address is translated to a hierarchical structured address for frame routing so that the number of forwarding table entries is reduced. However, it is possible that two MAC addresses are translated to the same structured address. Both [33] and [5] provided the concept of using cache to suppress broadcast traffic. By caching the most recently used dynamic directory entries at every PE node and using Relay PE maintaining the ARP entry, broadcasting messages can be reduced [33]. Similarly, Etherproxy in [5] caches the ARP entries it learned and suppresses broadcast messages it received by looking up the entries it cached. It retains the plug and play nature of Ethernet and is backward compatible. However, these cache schemes need extra memory to store the cache entries.

3. Caching on forwarding table (CFT)

Figure 1 shows a general Metro Ethernet including a provider network and multiple LAN segments. A provider network is composed of multiple Provider Edge (PE) nodes and Core Nodes (CNs) (switch or bridge). A LAN segment is composed of a Customer Edge (CE) node and multiple end users.

Ethernet nodes have the ability to dynamically learn the location of an end user by recording the port which the frame generated by the node comes from. Upon receiving a frame, a source address of the frame is learned in the forwarding table in an Ethernet node. A forwarding entry is in the format of $\langle MAC\ address, Port, Recordtime, age \rangle$. The subsequent frames destined to this address passing through the node could be forwarded to that port. ARP is used to solve an unknown MAC address for an IP address. Once a user broadcast an ARP request frame to a network, all the other users attached to it in the same VLAN will receive this frame. Only the destination user with the corresponding IP address replies back its MAC address. The ARP reply is sent via unicast. Both users record the other's IP and MAC addresses. The other users received the ARP message(s) also can learn the IP and MAC address mapping and record them to its ARP table.

To enable the caching effect in the forwarding table, CFT adds a corresponding IP address to a forwarding entry. The IP address can be learnt when a forwarding entry is created. Note that the source IP address is associated with the ARP request and reply frame. Now a forwarding entry

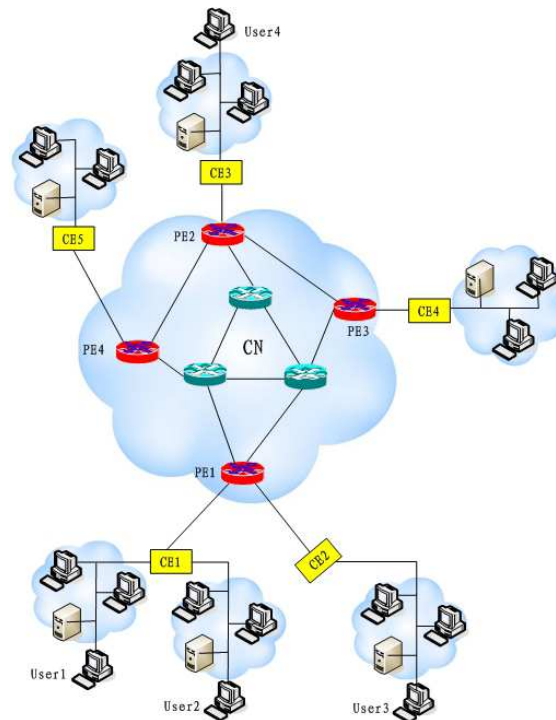


Figure 1. A Metro Ethernet

format in CFT is $\langle IP\ address, MAC\ address, port, recordtime, age \rangle$. Here IP is the end user's IP address; *recordtime* is the time when the entry is created or refreshed; and *age* indicates if an ARP entry is valid or not. A modified forwarding entry caches its ARP entry in the forwarding table, and hence can be used to answer ARP request. In this paper, we only deploy the CFT scheme in CE and PE nodes.

The address resolution and frame forwarding process are as follows. When a CE or PE node receives an ARP request frame, it first checks its forwarding table. If the requested IP address is found in the forwarding table, a reply frame is sent back. The broadcast is stopped at this node. Otherwise, the request frame is forwarded. If the forwarding entry of the source MAC address is not in the table, the node creates a forwarding entry. When a CE or PE node receives an ARP reply frame, it records/updates the corresponding entry in the forwarding table and forwards the frame. The flow chart in Figure 2 demonstrates how an ARP frame is processed.

Now let us use Figure 1 as an example. Assume user 1 intends to start a data session with user 4 whose ARP information has not been known by user 1. User 1 broadcasts an ARP request to the network to solve the MAC of user 4. When the broadcast frame reach CE 1, CE 1 uses the IP address of user 4 to search its forwarding table. The broadcast frame is stopped once a corresponding entry is found, and then a reply is sent to user 1. If there is no corresponding entry in the forwarding table, the broadcast frame is forwarded to PE 1 as well as the LAN segments user 2 belongs to. Similarly, after PE 1 receives the broadcast frame, it first searches its forwarding table. PE 1 broadcasts the request to the core network if no entry is matched. Otherwise, PE 1 stops the broadcast and replies an ARP reply back to user 1, while both CE 1 and PE 1 add the entry of user 1 or update the *recordtime* of user 1 in its forwarding table. Other PE and CE nodes transmit the broadcast frame process the frame as the same way as in legacy VLAN scheme. If the request frame reaches user 4, it just sends out a reply frame back. All the CE or PE nodes receive the request or reply frame learn the corresponding entry. Finally, user 1 could communicate with user 4.

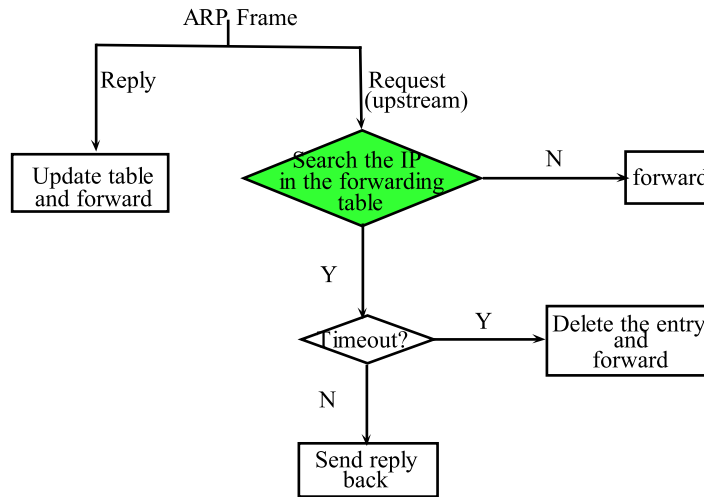


Figure 2. Process of a CE/PE handling an ARP frame

3.1. IP management in forwarding table

There are two ways to manage IP addresses. One way is to insert an IP address in each forwarding entry. Then an IP address is searched linearly when an ARP request comes. In this method, each entry only increases 4 bytes to store an IP address. A forwarding entry takes about 4/15 (4 bytes for IP address, 6 bytes for MAC address, 4 bytes for recordtime, 4 bytes for age and 1 byte for port number) more space. But the IP lookup time takes long. The second way is to divide the forwarding table into two parts. One part contains the index and the IP address which are sorted by the IP address. The other part is the original forwarding table by adding an index for each entry. The two parts are united by an index. A binary search can be used to find an unsolved IP address. In this method, at most $\lfloor \log_2(N) \rfloor + 1$ lookups are needed for an IP address matching. Here N is the number of IP addresses on the table. If the IP address has been found, the index is used to find the corresponding MAC address. The two tables are always updated at the same time. If the entry of a MAC address is outdated and deleted, the IP address with the same index will also be deleted. The table format is shown in Figure 3. This method can reduce the IP lookup time, but it costs two indexes space in the forwarding table. Assume that an index needs 2 bytes, then each entry needs 8/15 more spaces. Due to cache effect, a PE or a CE node handles less number of broadcast frames and hence the number of entries is reduced in CFT. The size of each entry is increased, but the total table is not increased as we will see in Section 6.

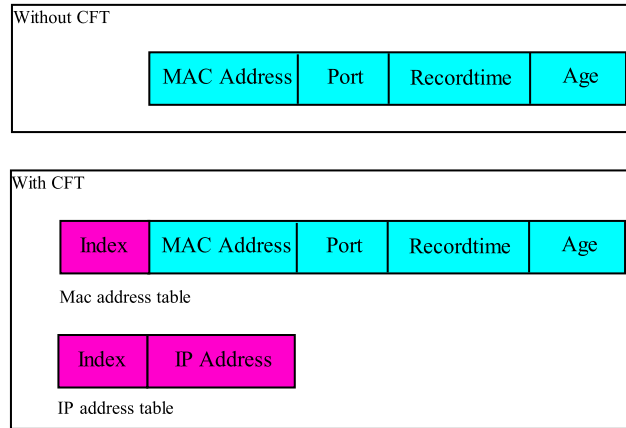


Figure 3. Forwarding Table Format

4. Cache in EMiM

The EMiM encapsulation scheme [34] can reduce the forwarding table sizes in PE nodes. In EMiM, the MiM encapsulation is done by the end user instead of the PE node. To allow an end user to do MiM encapsulation, the PE node's MAC address is associated with an ARP entry. The associated PE node's MAC (PMAC) address allows an end user to do MiM encapsulation, and hence a PE node does not need to maintain the entries of mapping end user's MAC address (EMAC) to PE node's MAC address, thus reducing the forwarding table size. The entry format of the ARP table is $\langle IP, EMAC, PMAC, recordtime, age \rangle$.

In the EMiM scheme, the forwarding table has no end user's entry, and hence the forwarding table cannot be used to cache ARP entry (CE node still can use CFT to cache ARP entries). But a PE node can still learn an ARP entry from an ARP request/reply frame. Hence we use a cache table to store ARP entries. A cache entry includes the *MAC address*, *IP address*, *PMAC address* and *recordtime*. The cache size is fixed. A Least Recently Used (LRU) replacement algorithm is used for entry management. This is because LRU can be easily implemented and an ARP entry is automatically timed out after some time. Besides LRU, a Modified Least Recently Used (MLRU) replacement algorithm is also introduced. In MLRU, the cache table is divided into two separate lists. List 1 is for entries that used more frequently, while list 2 is for entries with less frequency. When a new entry comes, it is stored in the head of list 2. When an entry is hit, it is put to the head of list 1. When list 1 is full, the entry in the tail is put to the head of list 2. If list 2 is full, the entry in the tail is deleted to release the space for new entries. Hence an entry has high hit ratio will stay longer than that in LRU, and hence MLRU can increase the hit ratio.

5. Hit ratio analysis

In this section, an analytical model is developed to examine the hit ratio in the CFT scheme at a PE node. The following assumptions are made in our model:

- The data access (to send frame) time of an end user follows Poisson distribution. The average data access time is T_a (one communication session per T_a seconds). R_i is the data access rate of user i .
- All the end users belong to one VLAN. Each user has probability to communicate with any other end users in the network.
- When an end user starts a data session, the destination is selected based on Zipf distribution with Zipf coefficient n .

- The forwarding tables are sufficient large to store all the entries and each forwarding table entry is timed out every T_o seconds.

Using Zipf's law, the probability of a user with rank i being selected is:

$$f(i, n) = (1/i^n) / \sum_{j=1}^N (1/j^n) \quad (1)$$

where N is the number of users;

The rate a node being selected as the destination by all the nodes under a PE node is:

$$\lambda_i = f_i * \sum_{j=1}^M R_j \quad (2)$$

here M is the number of users under the PE.

When the entry of node i comes to the forwarding table, it has T_o time to be stored there. If it is selected as a destination node in less than T_o time by a node under the PE, a hit is counted. Hence the hit ratio can be calculated as:

$$h_i = \int_0^{T_o} \lambda_i e^{-\lambda_i t} dt = 1 - e^{-\lambda_i T_o} \quad (3)$$

The total cache hit ratio for the PE node is:

$$h = \sum_{i=1}^N (f(i, n) h_i) \quad (4)$$

We compared the simulation results with the analytical results. The simulated Metro Ethernet have 4 PEs in the network. Every PE node has the same number of end users under it. We run 3600 seconds and the results of the last 2600 seconds are collected. The average data access rate is set to 180s. All the forwarding tables are set to be empty at start-up.

Figures 4 and 5 show the simulation and analytical modeling hit ratio with various value of N and R . From Figure 4, we know that the hit ratio increases as the increase of the number of end users. Figure 5 compares the results by varying the data access rate. The simulation with $n=1.0$ has 5 thousand end users while for $n=1.5$ has 6 thousand end users in the network.

The results show that the analytical results are well matched with the simulation results, particularly for $n=1.5$.

6. Performance evaluation

This section compares the performances of the proposed schemes and the legacy VLAN based scheme.

In this section, we simulated a Metro Ethernet with 50 thousand end users, 40 PEs, 150 CEs and 1000 VLANs in the network. Every PE node directly connects to at least 2 and at most 5 PE nodes. At least 2 and at most 6 CEs are behind a PE node. Every CE node can have up to 6 sites located under it. There are at least 8 and at most 256 end users behind a CE node. Each VLAN has at least 3 and at most 13 sites. Each user has probability to communicate with other end users in the same VLAN. The data access time follows Poisson distribution. The average data access time is T_a (one communication session per T_a seconds), and each data session lasts for average 20 seconds, randomly chosen between 1 and 39 seconds. When an end user starts a data session, it randomly picks a VLAN it belongs to, and the destination node is selected based on Zipf distribution. That means the probabilities of session destined to few addresses are heavier than others, which is more suitable for the real world traffic

situation [5]. The default forwarding table sizes are set to store at most 20,000 entries and each forwarding table entry is timed out every 120 seconds.

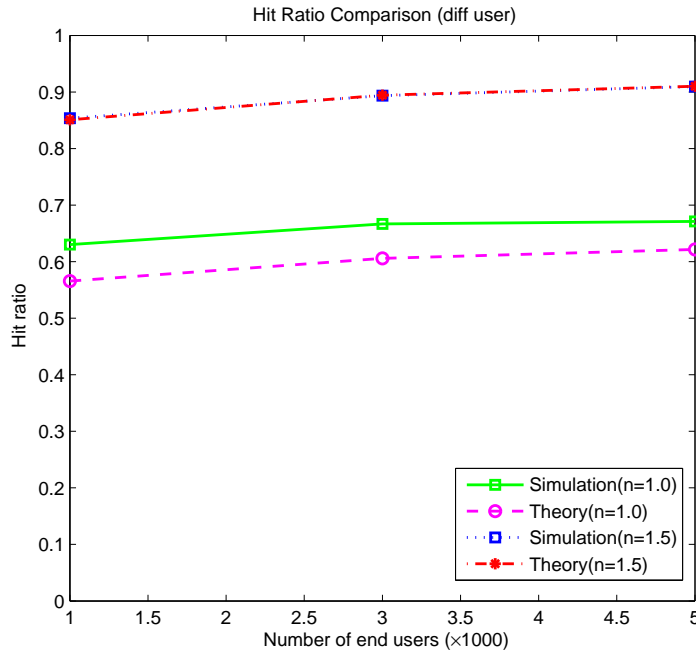


Figure 4. Hit Ratio Comparison (diff user)

In the simulation, we run 3600 seconds and the results after the system is steady are collected. All the forwarding tables are set to be empty at the beginning. We study the impacts of four parameters: session interval, Zipf coefficient, number of users and cache size (for the EMiM scheme only).

To evaluate the impact of session interval, we varied the session interval from 60s to 240s in case 1. The data access rate is set to 60s for case 2 and case 4 and 120s for case 3. In the simulation, the PE messages include the data messages PE handles and the broadcast/multicast messages a PE generates to resolve unknown MAC addresses. User messages contain all the data messages and the broadcast/multicast messages users generated to resolve unknown destination MAC addresses.

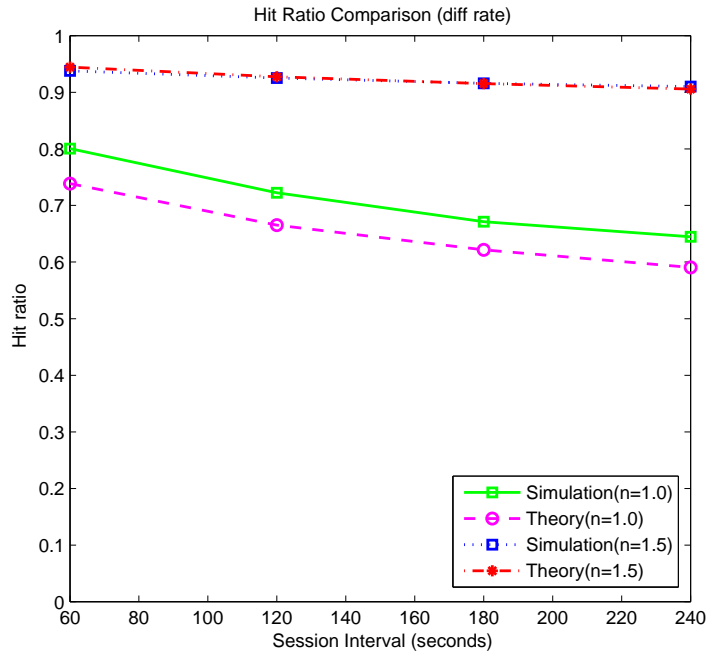


Figure 5. Hit Ratio Comparison (diff rate)

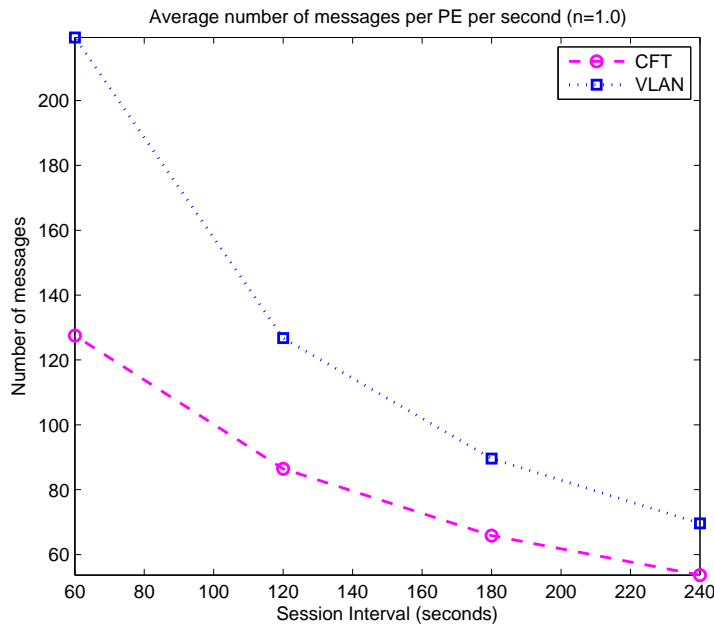


Figure 6. Average PE message per second

Case 1: Impact of session interval

In this case, T varies from 60 to 240 seconds. Zipf coefficient n is set to 1.0. Figures 6 and 7 demonstrate the average number of messages a user handled per second and the average number of messages a PE generated per second. From the Figures, we can see that the CFT

scheme can significantly reduce the messages handled by PEs or end users compared to the legacy VLAN scheme, particularly in heavier traffic case. Note that session interval is the reciprocal of data access rate. As the data access rate increases, the number of message grows more rapidly in legacy VLAN system. This is due to more broadcast frames needed to resolve unknown MAC addresses. But in the CFT scheme, more traffic results in high cache hit ratio and hence reduce more broadcast messages. Hence the number of messages increases much less than that in legacy VLAN scheme. At the session interval equals to 60s, around 50% PE messages can be reduced for CFT per PE, and about 30% messages can be reduced for each end user.

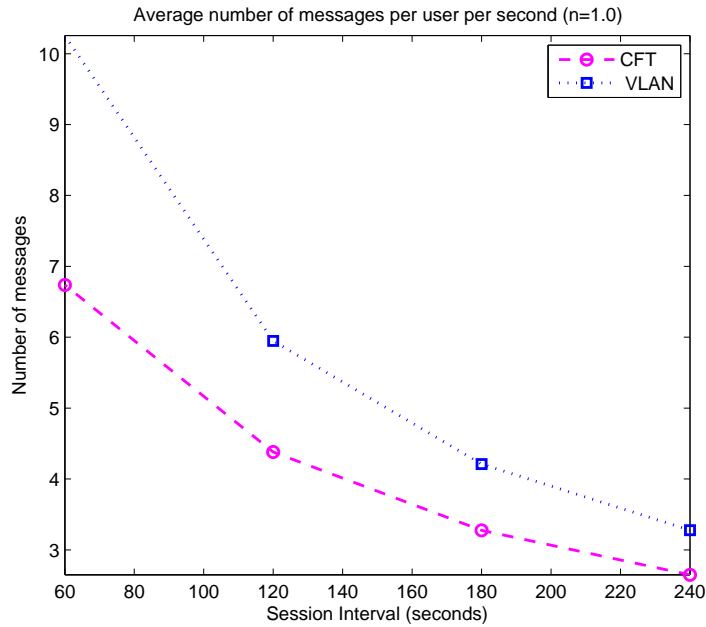


Figure 7. Average User message per second

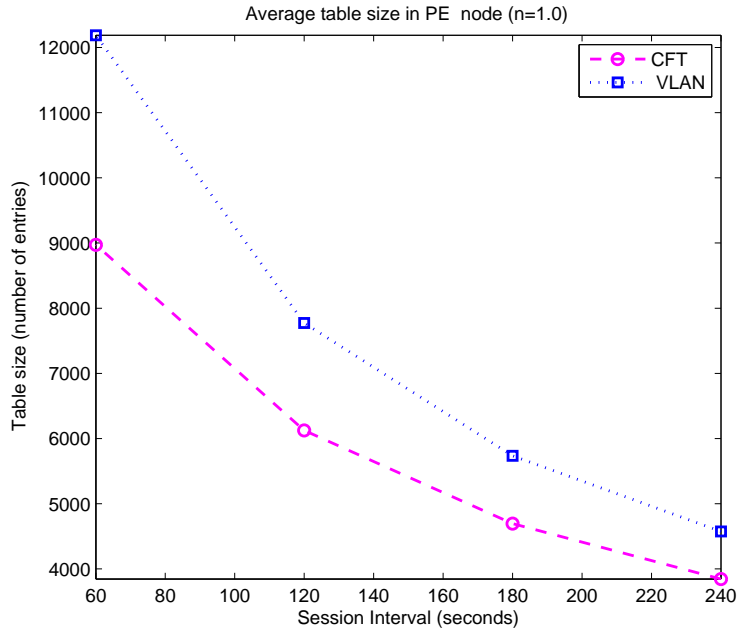


Figure 8. Average PE table size

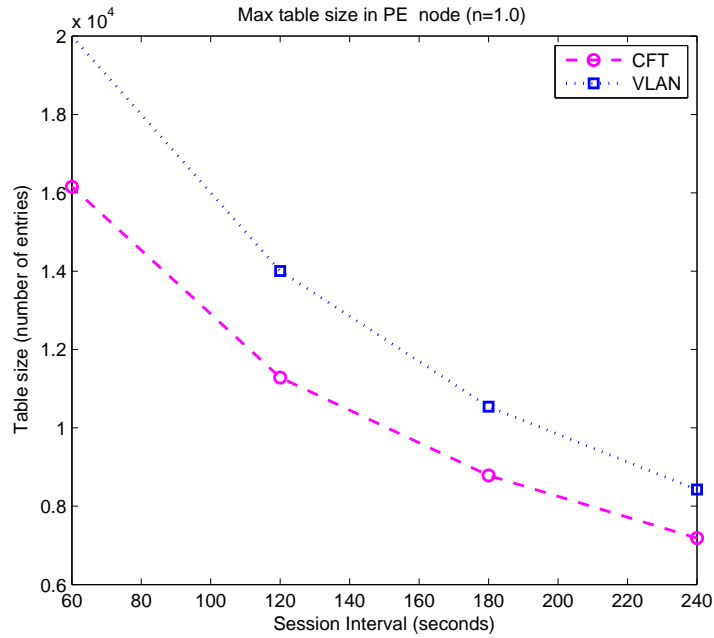


Figure 9. Maximum PE table size

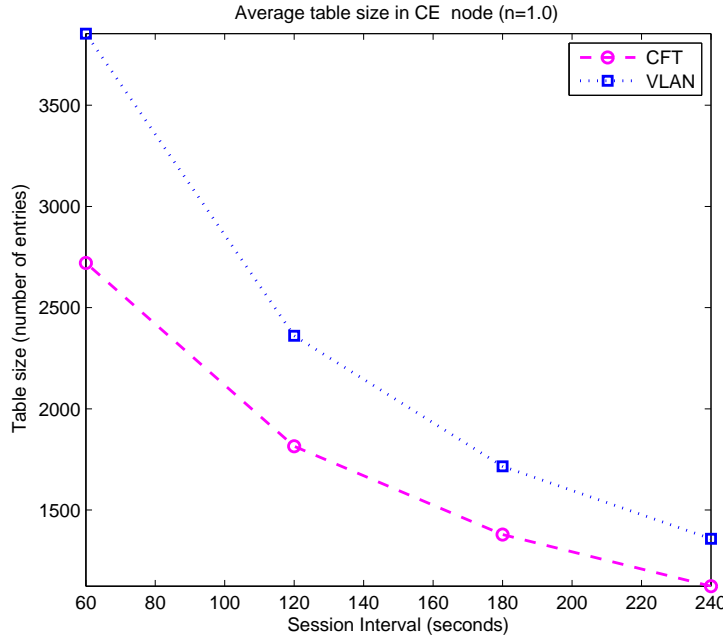


Figure 10. Average CE table size

Figures 8 and 9 illustrate the average table size and the maximum table size in a PE. Figure 10 and Figure 11 show the average table size and the maximum table size in a CE. Both the average and maximum table sizes increase as the decrease of session interval. This is due to shorter session interval results in high broadcast and hence more forwarding entries. The results clearly show that the CFT scheme can reduce both average and maximum table size in PE and CE tables. As we pointed out that the CFT scheme needs more space for each forwarding entry, but the maximum table size is reduced and hence the total memory space is also equal.

The hit ratio of CFT is plotted in Figure 12. The hit ratio is computed as the number of ARP requests answered by the cache divided by the number of ARP requests received by a PE or CE. The result shows that the hit ratio is reversely proportional to the session interval time. This is due to less session interval time corresponding to high broadcast requests, and hence high probability to be answered by cache. When the data interval time is at 60 seconds, the hit ratio reaches about 45%, this can significantly reduce the broadcast messages.

Case 2: Impact of Zipf coefficient

In this section, we vary the Zipf coefficient from 0.5 to 1.5. A high Zipf coefficient stands for the destination users are more focused on the popular users.

The average message number a PE (a user) handled per second are demonstrated in Figures 13 and 14. The CFT scheme has better performance for larger Zipf coefficient. This is because the destination users are more focused on the popular ones for a larger coefficient system than that for a smaller coefficient system. At 1.5 Zipf, more than 60% PE messages can be reduced for average PE.

Figure 15 presents the maximum PE table size in the CFT scheme and the VLAN scheme. Similarly, the maximum PE table size is reduced as the Zipf coefficient increases. A larger Zipf coefficient results in fewer destination users and hence fewer number of forwarding entries in the forwarding table. When Zipf coefficient is 1.5, the maximum table size is reduced about 40% in CFT.

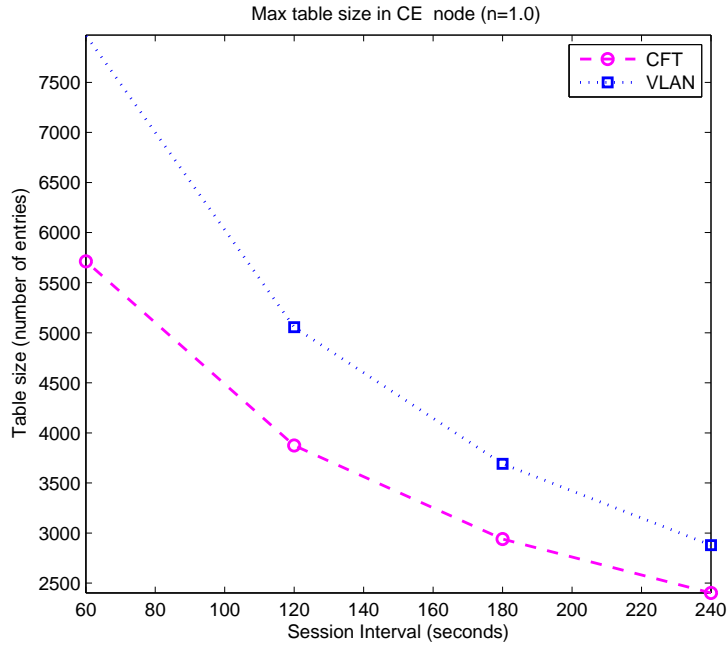


Figure 11. Maximum CE table size

Figure 16 shows the hit ratio varying with the Zipf coefficient. When the Zipf coefficient increases, the chance users start sessions with the same destination increase. And hence the hit ratio increases. At 1.5 Zipf coefficient, the hit ratio can be over 60%. Hence the CFT scheme is more efficient for a system with some popular users.

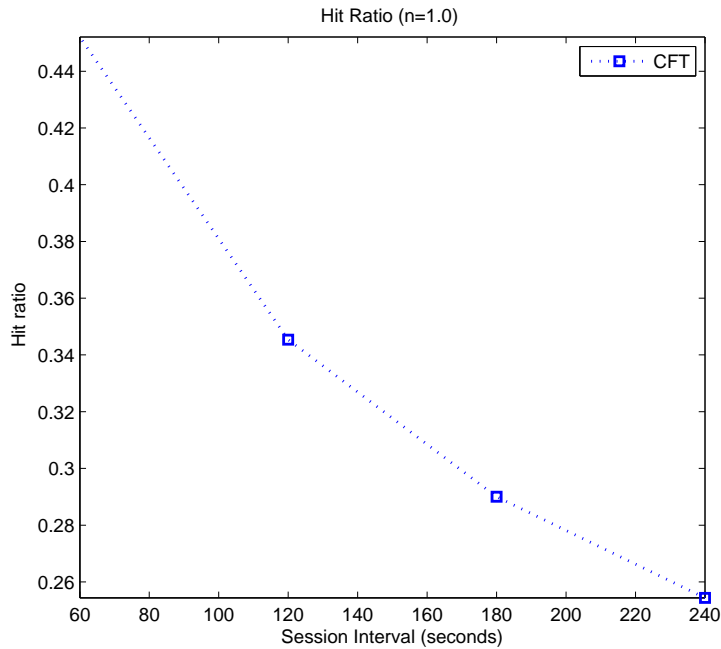


Figure 12. Hit Ratio

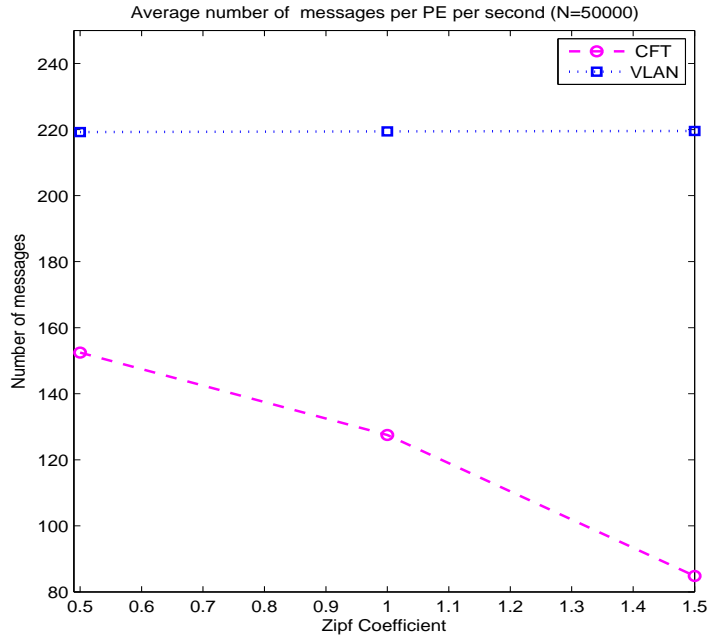


Figure 13. Average PE message per second

Case 3: Impact of number of users

In this case, we vary the number of users from 10,000 to 35,000 to study the impact of the number of users. The data access rate is set to 120 seconds and the Zipf coefficient is set to 1.0. We compare the performance of the CFT scheme with cache in EMiM and the VLAN scheme. The cache size in EMiM is large enough to record all the entries.

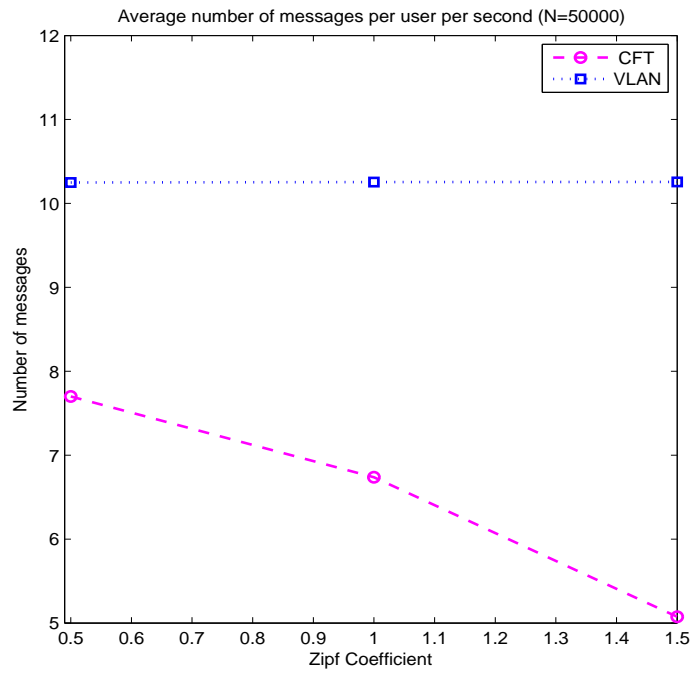


Figure 14. Average User message per second

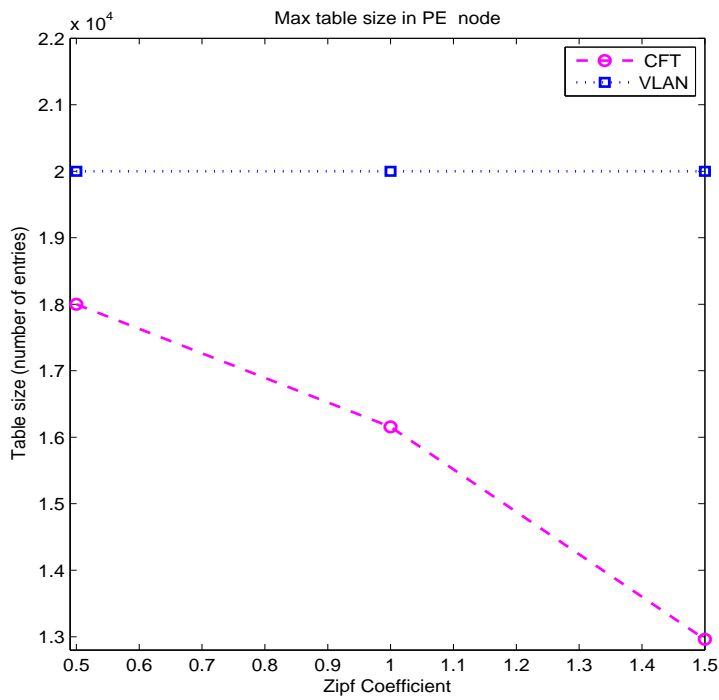


Figure 15. Maximum PE table size

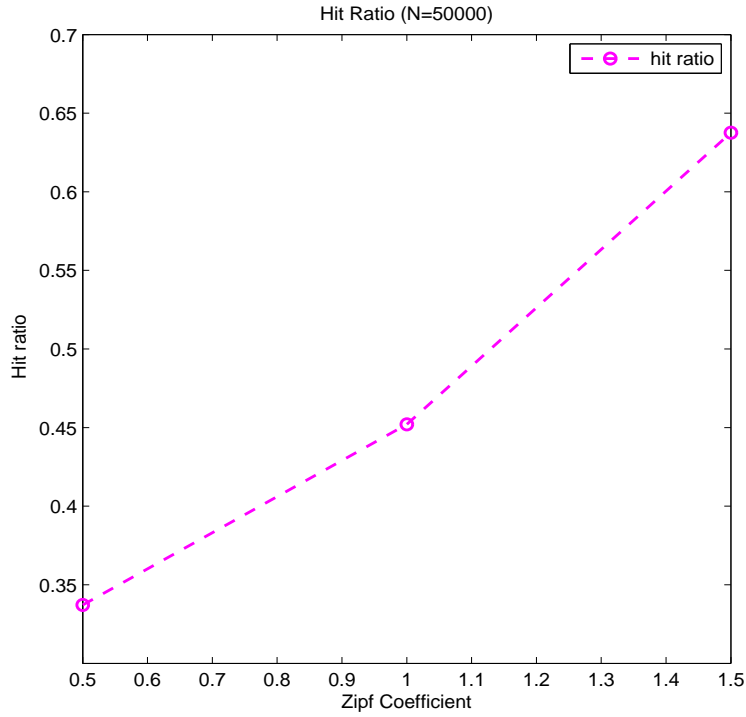


Figure 16. Hit Ratio

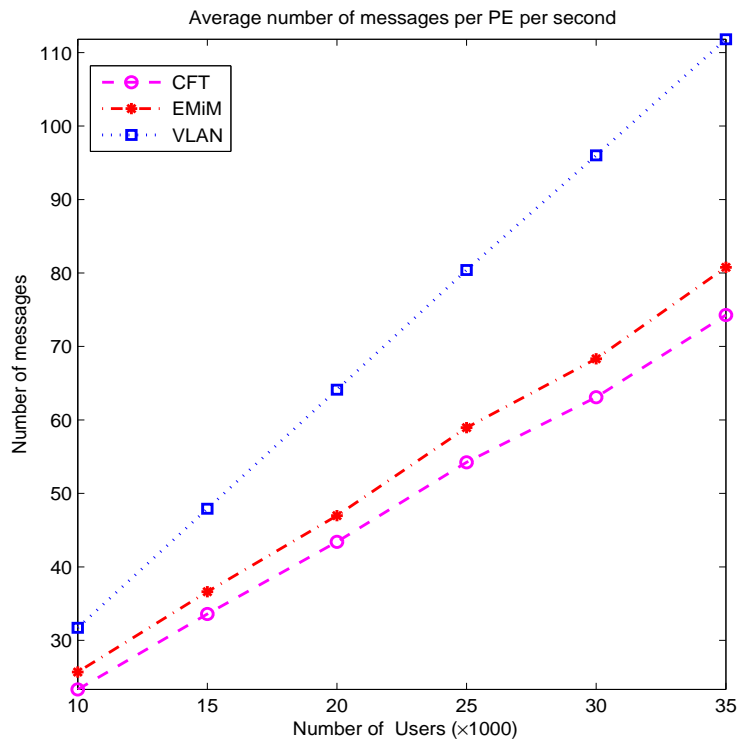


Figure 17. Average PE message per second

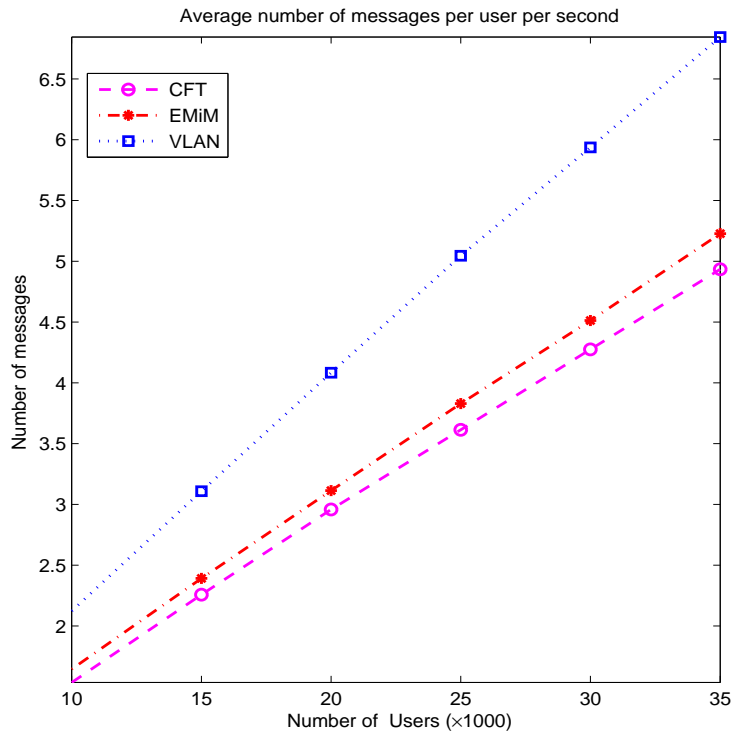


Figure 18. Average User message per second

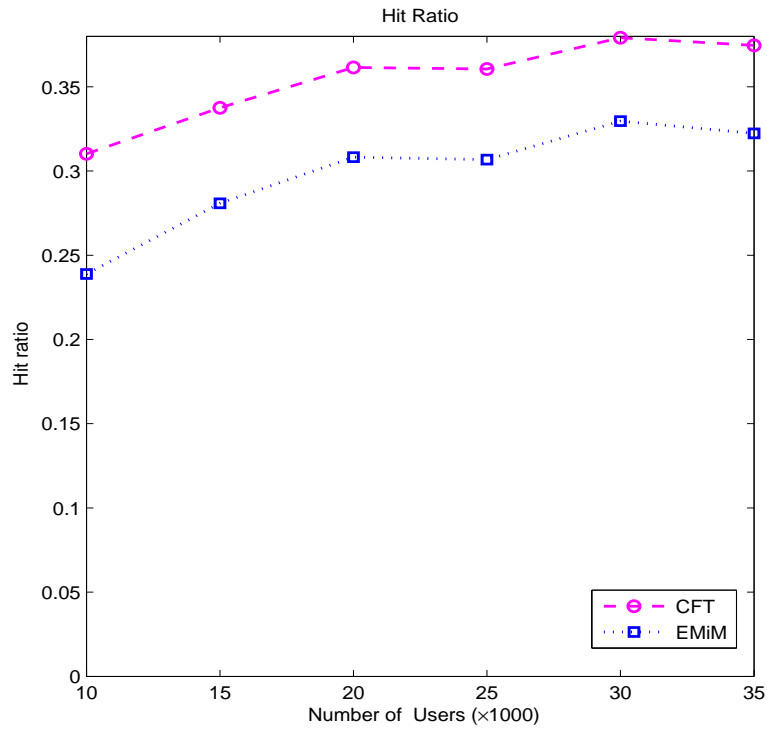


Figure 19. Hit Ratio

Figures 17 and 18 show the number of messages a PE node and an end user handled per second. The number of messages increases as the number of users increases. As the number of users increases, the number of broadcasts increases and hence a PE node and a user need to handle more messages. The number of messages for all the three schemes increases linearly. But the VLAN scheme has larger slope, i.e., increase rate higher than the other two schemes. Caching in EMiM has similar slope as that of the CFT scheme, but a little bit more messages than that of the CFT scheme.

From Figure 19, we observe that the hit ratio increases along with the increase of the number of users when the number of users is less than 30,000. However, when the number of users reaches 30,000, the hit ratio is almost a constant by further increase the number of users. When the number of users is small, increasing the number of users can enlarge the probability of the most popular users to be selected as the destination nodes, and hence increase the hit ratio. But when the number of users is over a threshold, more and more unpopular users have been selected and hence make the hit ratio nearly a constant.

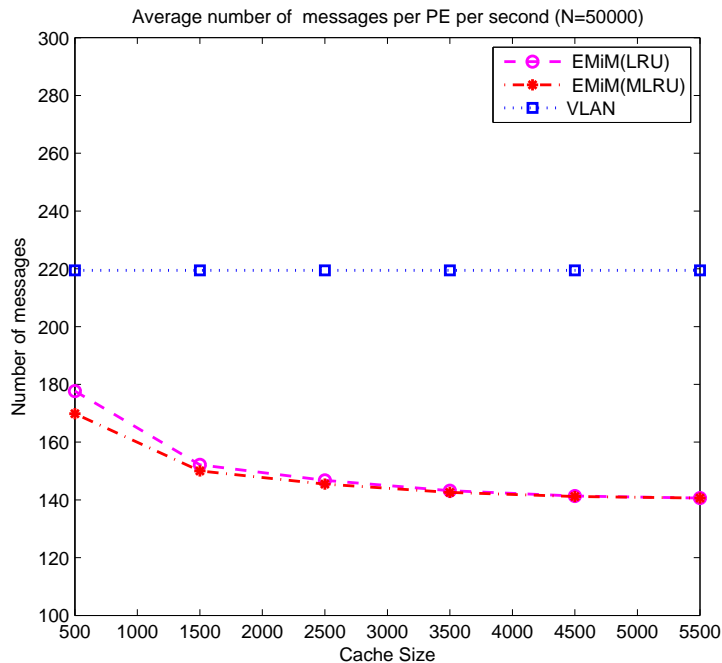


Figure 20. Average PE message per second

Case 4: Impact of cache size

This case studies the impact of cache size in the EMiM scheme. In this experiment, the network has 50,000 users and the Zipf coefficient is set to 1.0. The performance of LRU and MLRU entry replacement algorithm are compared with the VLAN scheme.

Figure 20 illustrates the average message numbers handled by a PE node per second. The number of messages decreases as the cache size increases. When cache size increases from 500 to 1500 entries, more than 12% message can be reduced compared with that in the VLAN scheme; But when the cache size increases from 1500 to 2500 entries, only less than 3% more messages could be reduced. The results show that the most hits are contributed from the most popular ones. Further increasing the cache size only stores more one time entries, and these entries never be hit again in its valid time, and hence we only need to keep a relative small

cache in PE nodes.

Figure 21 demonstrates the hit ratio in a PE cache. From the figure, we know that increasing the cache size is a very effective way to improve the hit ratio when the cache is small. After the cache size reaches a value (here it is 40,000 entries), further increasing the size only has little improvement. This shows that the cache hit ratio is obtained from the most popular entries, storing more unpopular ones has very little effect to improve performance. MLRU algorithm performs slightly better than LRU algorithm especially when the cache size is small. This is because the more frequently used entries could be stayed longer in MLRU scheme.

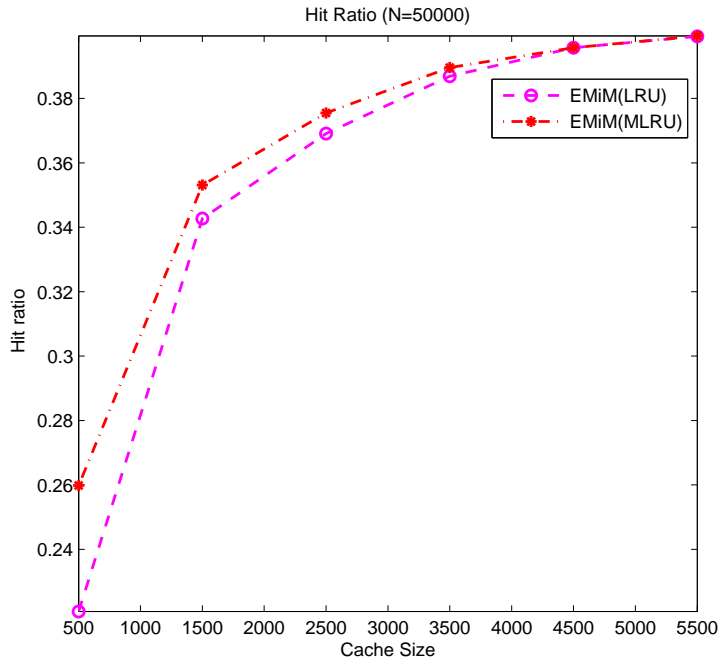


Figure 21. Hit Ratio

7. Conclusions and future work

Highly development of Ethernet technologies has made them attractive as a MAN transport for service providers [20]. However, Ethernet is originally designed for LAN area which handles limited number of users. When deploy Ethernet in MAN area, some loss in efficiency and scalability has come out. Specially, by using broadcast-based mechanism to resolve the location of an unknown address.

This paper proposes an efficient cache scheme for Metro Ethernet, called Cache in Forwarding Table (CFT). CFT learns the IP and MAC mapping in a frame and eliminates the subsequent broadcast frames whenever a request is answered by cached entry. We also discussed with cache in EMiM. LRU and MLRU cache replacement are compared. The simulation results show the proposed scheme can save broadcast messages for address resolution and reduces the forwarding table size in PE nodes. Moreover, it inherits the plug-and-play setup and self-configuration nature of Ethernet.

ACKNOWLEDGEMENT

This work is supported by the RGC of Hong Kong Polytechnic University under the Project

Code A-PJ69.

References

- [1] IEEE Std 802.3x-1997 and IEEE Std 802.3y-1997.
- [2] J. Postel and J. Reynolds, "A Standard for the Transmission of IP Datagrams over IEEE 802 Networks", *RFC 1042*, 1988.
- [3] M. Casado, M. J. Freedman and S. Shenker, "Ethere: Taking Control of the Enterprise", *ACM SIGCOM*, 2007.
- [4] G. Chiruvolu, "Issues and Approaches on Extending Ethernet Beyond LANs", *IEEE Communication Magazine*, v42(3), pp 80-86, 2004.
- [5] K. Elmeleegy and A. L. Cox, "EtherProxy: Scaling Ethernet By Suppressing Broadcast Traffic", *The 28th Conference on Computer Communications. IEEE*, 2009.
- [6] A. Myers, T.E. Ng, and H. Zhang, "Rethinking the Service Model: Scaling Ethernet to a Million Nodes", *Third Workshop on Hot Topics in Networks (HotNets-III)*, 2004.
- [7] D.C. Plummer, "An Ethernet Address Resolution Protocol or Converting Network Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", *RFC 826*, 1982.
- [8] R. Droms, "Dynamic host configuration protocol", *RFC 2131*, 1997.
- [9] S. Halabi, *Metro Ethernet*, Cisco Press, 2003.
- [10] "Problems with broadcasts", http://www.ists.dartmouth.edu/classroom/crs/arp_broadcast.php.
- [11] IEEE Std. 802.1Q, "Virtual Bridged Local Area Networks".
- [12] T. L. Rodeheffer, C. A. Thekkath and D. C. Anderson, "SmartBridge: a Scalable Bridge Architecture", *ACM SIGCOMM*, 2000.
- [13] IEEE 802.1ah, "Provider Backbone Bridges".
- [14] R. Perlman, "Rbridges: Transparent routing". *IEEE Infocom*, 2004.
- [15] R. Garcia, J. Duato, and F. Silla, "LSOM: A Link State Protocol Over Mac Addresses for Metropolitan Backbones Using Optical Ethernet Switches", *The Second IEEE International Symposium on Network Computing and Applications (NCA)*, 2003.
- [16] IEEE 802.1D-2004, "Spanning Tree Protocol revision of IEEE std 802.1D".
- [17] IEEE 802.1S, "Multiple Spanning Tree".
- [18] IEEE 802.1W, "Rapid Reconfiguration of Spanning Tree".
- [19] Huynh Minh, Mohapatra Prasant, "A Scalable Hybrid Approach to Switching in Metro Ethernet Networks Local Computer Networks", *The 32nd IEEE Conference on Local Computer Networks*, 2007.
- [20] Ibanez G., Garcia A., Azcorra A., "Alternative multiple spanning tree protocol (AMSTP) for optical Ethernet backbones" *The 29th Annual IEEE International Conference on Local Computer Networks*, 2004.
- [21] "IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks", *IEEE*, 1998.
- [22] M. Ali, G. Chiruvolu and A. Ge, "Traffic Engineering in Metro Ethernet", *IEEE Networks*, v19(2), pp10-17, 2005.
- [23] IEEE 802.1Q, "Virtual LANs".
- [24] S. Ray, R. A. Guerin and R. Sofia, "A Distributed Hash Table Based Address Resolution Scheme for Large-Scale Ethernet Networks", *IEEE ICC*, 2007.
- [25] Changhoon Kim, Rexford, J., "Revisiting Ethernet: Plug-and-play made scalable and efficient", *The 15th IEEE Workshop on Local and Metropolitan Area Networks*, 2007.
- [26] Changhoon Kim, Matthew Caesar, and Jennifer Rexford, "Floodless in SEATTLE: A Scalable Ethernet Architecture for Large Enterprises", *ACM SIGCOMM*, 2008.
- [27] P. Wang, C. Chan, and P. Lin, "Translation for Enabling Scalable Virtual Private LAN Service", *21st International Conference on Advanced Information Networking and Applications Workshops*, 2007.
- [28] "BellSouth Metro Ethernet", <http://www.bellsouthlargebusiness.com>.
- [29] Huynh Minh, Mohapatra Prasant, Goose Stuart, "Cross-over spanning trees Enhancing metro Ethernet resilience and load balancing", *BROADNETS*, 2007.
- [30] Minh Huynh, Prasant Mohapatra, Stuart Goose, "Spanning tree elevation protocol: Enhancing metro Ethernet performance and QoS", *Computer Communications*, 2009.
- [31] Mirjalily G., Karimi M.H., Adibnia F., Rajai S, "An Approach to Select the Best Spanning Tree in Metro Ethernet Networks", *Computer and Information Technology*, 2008.
- [32] M.Tafti, G. Mirjalily, and S. Rajae, "Topology Design of Metro Ethernet Networks Based on Load Balance Criterion", *International Symposium on Telecommunications*, 2008.
- [33] Ravi Kumar Buregoni, "A Unified Distributed Directory based Service Delivery Architecture for Metro Ethernet Networks", *11th International Conference on Advanced Communication Technology*, 2009.
- [34] X. Sun, Z. Wang, H. Che, and F. Zhao, "An End User Enabled MAC-in-MAC Encapsulation Scheme for Metro Ethernet", *Proceedings of the 2008 IEEE International Symposium on Parallel and Distributed Processing with Applications*, 2008.