산업융합연구 제17권 제4호 pp. 131-137, 2019
ISSN 2635-8875 / e-ISSN 2672-0124
https://doi.org/10.22678/JIC.2019.17.4.131

User Centric Cache Allocation Schemes in Infrastructure Wireless Mesh Networks

Seung Hyun Jeon
Researcher, School of Electrical Engineering, Korea Advanced Institute of Science and Technology

인프라스트럭처 무선 메쉬 네트워크에서 사용자 중심 캐싱 할당 기법

전승현
한국과학기술원 전기및전자공학부 박사후 연구원

요 약 전통적으로 예상 전송 시간 메트릭을 라우팅으로 사용하는 인프라스트럭처 무선 메쉬 네트워크는 모바일 유저에게 심각한 지연을 발생한다. 캐시를 갖는 메쉬 라우터를 통해 성능 향상을 기대할 수 있지만, 콘텐츠 인기도, 캐시 비용 및 다운로드 지연을 고려한 캐싱 할당 기법에 대한 연구는 부족하다. 이러한 문제점을 해결하고자 모바일 유저 수의 증가에 따라 성능을 극대화할 수 있는 최적화 기반 사용자 중심의 캐싱 기법을 제안한다. 또한 메타 휴리스틱 기법인 유전 알고리즘 (일정 교차/변이)을 적용하여 예상 지연 시간을 줄이며 캐시 적중률에 만족하는 최적의 모바일 유저 수를 찾는다. 이는 인프라스트럭처 무선 메쉬 네트워크를 운영하는 통신 사업자가 다양한 캐싱 비용, 라우터 속도, 콘텐츠 인기도를 고려한 목표 캐시 적중률을 만족할 뿐만 아니라 성능 향상을 기대할 수 있다.

키워드 : 캐시 할당, 인프라스트럭처 무선 메쉬 네트워크, 콘텐츠 인기도, 캐시 비용, 유전 알고리즘, 최적화

Abstract In infrastructure wireless mesh networks (WMNs), in order to improve mobile users' satisfaction for the given cache hit ratio, we investigate an User centric Cache Allocation (UCA) scheme while reducing cache cost in a mesh router (MR) and expected transmission time (ETT) for content search in cache. To minimize ETT values of mobile users, a genetic algorithm based UCA (GA-UCA) scheme is provided. The goal is to maximize mobile users' satisfaction via our well defined utility, which considers content popularity and the number of mobile users. Finally, through solving optimization problem we show the optimal cache can be allocated for UCA and GA-UCA. Besides, a WMN provider can find the optimal number of mobile users for user centric cache allocation in infrastructure WMNs.

Key Words : Cache Allocation, Infrastructure WMNs, Content Popularity, Cache Cost, Genetic Algorithm, Convex Optimization

1. Introduction

For a long time, infrastructure wireless mesh networks (WMNs) have been studied[1-3]. Infrastructure WMNs consist of WiFi access points and they can be connected by multi-hops in a campus or an office. However, considering expected transmission time (ETT) as the most popular routing metric in WMNs, throughput is dramatically dropped by multi-hop routing[4]. Moreover, when mobile users download popular content from video on demand servers or

Corresponding Author : Seung Hyun Jeon(creemur@kaist.ac.kr)
Received December 10, 2019 Revised December 14, 2019
Accepted December 20, 2019 Published December 31, 2019

peer-to-peer (P2P) content sharing, severe downloading delay can occur due to collisions among mobile users[3]. Thus, for solving these problems cache usage in WMNs has been induced.

A mesh router (MR) as gateway in infrastructure WMNs can store content in cache[3,5]. Alasaad *et al.* improved the performance of throughput using content replication strategy in infrastructure WMNs[3]. Also, Das *et al.* presented MechCache, which is a transparent caching system to mitigate bottleneck effect at the gateway MR[5]. Accordingly, the use of cache in a MR can reduce downloading delay for popular content and mitigate collision among mobile users. Besides, as increasing cache cost including software, hardware, and operation cost can reduce profits of cellular providers[6], WMN providers should consider cost on cache allocation to MRs. Generally, cache allocation is required by considering content popularity based on Zipf-like distribution[7]. The existing cache allocation solutions dependent on content placement can locate cache near the network edge[8] or at content-centric networking (CCN) routers[9]. However, it is challenging to store content to cache of numerous MRs in infrastructure WMNs.

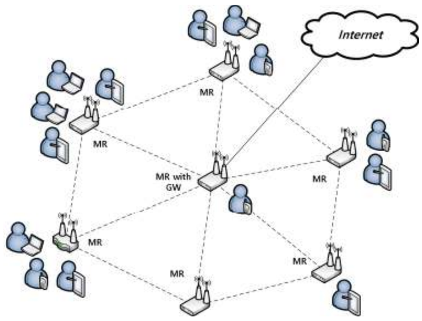
Here, we tackle cache allocation problem in infrastructure WMNs. Since for high cache hit ratio, redundant content can increase at cache of MRs in WMNs, a WMN provider may spend considerable expense on increase of cache usage. On the contract, for low cache hit ratio, if the content to be watched does not exist in cache of the MR to be first accessed, additional ETT for content search may affect mobile users' downloading delay. Moreover, considering user traffic corresponding to increased mobile users and content popularity, an user centric cache allocation scheme for content in infrastructure WMNs is required.

Zhu *et al.* presented unified P2P and cache framework in infrastructure WMNs[2]. Video content at the selected MRs is cached and ETT is also considered as routing metric. However, they do not consider cache allocation based on the content popularity. Wang *et al.* discussed on the cache allocation problem for CCN via content placement optimization[9]. The factors on the optimal cache allocation such as network topology and content popularity were investigated. However, cache allocation in WMNs should consider routing metric such as ETT. Vakali *et al.* introduced Web cache content replacement using genetic algorithm (GA)[10]. Through GA based cache replacement, cache hit ratio was improved, compared with the conventional least-recently-used (LRU). However, we improve mobile users' satisfaction by reducing ETT via cache relocation for content using GA. Thus, via previous work there is lack of studies on cache allocation considering mobile users' satisfaction, provider's cache cost, and downloading delay in infrastructure WMNs.

In this paper, we propose user centric cache allocation schemes to assign amount of the optimal cache each content in a MR, by recognizing the user characteristic such as content popularity and the number of increased mobile users. The goal of the proposed schemes is to maximize mobile users' satisfaction via user centric cache allocation. Here, the term of *user centric* represents the recognition of user profile and satisfaction in a proposed utility function which reflects characteristic of content for cache hit ratio, content popularity, and the number of mobile users based on proportional fairness[11]. Finally, we find user centric cache allocation schemes between mobile users and a WMN provider using convex optimization[12] and GA[10].

2. Problem Formulation

2.1 System Model



[Fig. 1] Infrastructure WMNs

We address cache allocation schemes for infrastructure WMNs, as in Fig. 1. We assume all MRs are connected in wireless multi-hop networks to reduce installation cost of wireline links for WMNs. We define the content popularity[7]. Based on this, mobile users in order to watch the content can access to the MR within the nearest distance. Mobile users are randomly positioned within 200×200 meters. Via one of MRs, other MRs can connect to the Internet. We assume MRs have IEEE 802.11n radio frequency and are deployed in an outdoor environment. Besides, MRs exchange content information within their cache periodically[13]. If all mobile users can not find content at cache of the MR to be first accessed, they can search the content by the routing of the MR.

2.2 Proposed User Centric Cache Allocation Scheme

We present an User centric Cache Allocation (UCA) scheme for infrastructure WMNs. A proposed utility function is defined as following:

$$\begin{aligned} U = \max_{s^k} & \sum_{j=1}^{N_U}\sum_{k=1}^{K}\log\left\{1+\eta H^k(A_k)p_j^k s^k\right\} \\ & -\sum_{k=1}^{K}\sum_{i=1}^{N_{MR}} s^k \alpha_{ki} c_{ca} - \sum_{j=1}^{N_U}\sum_{k=1}^{K} ETT^k(A_k) \\ & s.t. s^k \geq 0, A_k = (\alpha_{ki} \in \{0,1\}), \end{aligned} \quad (1)$$

where η is scalable factor, Ak is chromosome vector on the existence for k th content in MRs, pkj is k th content popularity from j th mobile user's request, sk is cache allocation for k th content in a MR, and cca is cost for storing content in cache per MR paid from a WMN provider (e.g., operation cost). Here, using Ak, cache hit ratio for k th content is defined as Hk(Ak) and ETT for k th content is defined as ETTk(Ak). Besides, αki is defined as gene on the existence of k th content within i th MR.

The proposed utility consists of three terms. The first term represents all mobile users' satisfaction from cache allocation for k th content (sk). This term originates from the concept of proportionally fair in WMNs[11] because mobile users fairly contend for amount of cache assigned in order to download popularity based content from a MR. The second term represents cache cost per MR paid from a WMN provider. Actually, cache allocation for k th content can share all mobile users. However, since we consider cache usage as operation cost, we assume the cache cost should be paid from a WMN provider for high satisfaction to mobile users. Finally, the third term represents all mobile users' ETT dependent on Ak. The second and third terms affect mobile users' satisfaction. Hence, we define them as negative equations.

The goal of the proposed UCA scheme is to find the optimal cache allocation in order to maximize the total utility that sums mobile users' satisfaction, subtracts the total assigned cache cost of the WMN provider, and subtracts mobile users' delay for downloading content.

For the first term, Hk(Ak) is defined as

$$H^k(A_k) = \frac{1}{N_{MR}} \times \sum_{i=1}^{N_{MR}} \alpha_{ki} \leq 1, \quad (2)$$

where NMR is the number of MRs in WMNs. If all αki equal to 1, cache hit ratio for k th content is set as Hk(Ak)=1.

For the third term, ETTk(Ak) is divided as followings due to having two paths in infrastructure WMNs:

$$ETT_k(A_k)=\begin{cases}\sum_{i=1}^{N_U^1}\alpha_{ki}ETT^k\ as\, 1-hop,\\ \sum_{i=1}^{N_U^2}(1-\alpha_{ki})(ETT^k)^2\ as\, 2-hop,\end{cases} \qquad (3)$$

where N1U and N2U are defined as the number of mobile users with 1-hop's and 2-hop's delay required for content search, respectively, and $ETT_k=\frac{1}{1-p}\times\frac{s^k}{B}$ [14]. When cache of the MR to be first accessed has the content, the mobile users belong to N1U. Otherwise, the others belong to N2U. Let p denote the probability that the transmission (i.e., copy) of the k th content within the cache of i th MR is not successful. That is, $p=1-(1-p_c)(1-p_e)$. p_c and p_e are collision and error probability [15]. Let B denote the link bandwidth.

We can solve the optimization problem on two cases for each ETT[12]. First, the optimization problem for 1-hop can be derived as follows:

$$\frac{\partial^2 U}{\partial^2 s^k}=-\sum_{j=1}^{N_U^1}\sum_{k=1}^{K}\frac{\{\eta H^k(A_k)p_j^k\}^2}{\{1+\eta H^k(A_k)p_j^k s^k\}^2}<0. \qquad (8)$$

Therefore, in case of mobile users with 1-hop delay for content search, optimal cache allocation for the k th content in each MR can be derive as following:

$$(s^k)^*=\sum_{k=1}^{K}\sum_{j=1}^{N_U^1}\frac{\eta H^k(A_k)p_j^k}{1+\eta H^k(A_k)p_j^k}\times\left\{\sum_{k=1}^{K}\sum_{i=1}^{N_{MR}}\left(\alpha_{ki}c_{ca}\sum_{j=1}^{N_U^1}\frac{\alpha_{ki}}{(1-p)B}\right)\right\}^{-1}.$$

(9)

On the other hand, the optimization problem for 1-hop[12] can be derived as follows:

$$\frac{\partial^2 U}{\partial^2 s^k}=-\sum_{j=1}^{N_U^2}\sum_{k=1}^{K}\frac{(\eta H^k(A_k)p_j^k)^2}{(1+\eta H^k(A_k)p_j^k s^k)^2}-\sum_{j=1}^{N_U^2}\sum_{k=1}^{K}\sum_{i=1}^{N_{MR}}\frac{2(1-\alpha_{ki})}{((1-p)B)^2}<0.$$

(10)

However, in case of mobile users with 1-hop delay for content search, optimal cache allocation for k th content in WMNs is derived as a quadratic equation like following: $a(s^k)^2+bs^k+c=0$. Therefore, the optimal cache allocation can be derive as following:

$$(s^k)^*=\frac{-b+\sqrt{b^2-4ac}}{2a}\ge 0, \qquad (11)$$

where $a=\sum_{k=1}^{K}\sum_{i=1}^{N_{MR}}\sum_{j=1}^{N_U^2}\frac{2(1-\alpha_{ki})}{((1-p)B)^2}, b=\sum_{k=1}^{K}\sum_{i=1}^{N_{MR}}\alpha_{ki}c_{ca}$, and $c=-\sum_{k=1}^{K}\sum_{j=1}^{N_U^2}\frac{\eta H^k(A_k)p_j^k}{1+\eta H^k(A_k)p_j^k}$. Therefore, through (7), (8), and (10), since the proposed utility (1) can be a concave function, it can guarantee the existence of a global optimal point[12].

2.3 Proposed Genetic Algorithm based User Centric Cache Allocation Scheme

We present a GA based User centric Cache Allocation (GA-UCA) scheme for infrastructure WMNs. By changing the location of k th content into cache of other MRs using GA technology, ETT values of mobile users can be updated. That is, using GA-UCA k th content that mobile users watch is transferred from the 2-hop MP to the 1-hop MR (i.e., a near MR). However, for copying the content to other MRs, it takes the transmission time. Accordingly, we assume k th content can be shifted between cache and storage at the MR on requesting the GA-UCA scheme.

For Hk(Ak) of the UCA scheme, we apply uniform crossover and mutation of GA in Fig. 2.

```
Input : generation iteration, K, H^k(A_k), N_MR as R
Output : α_k^*
Population ← InitializePoplation(generation iteration, K)
EvaluatePopulation(Population)
α_k^* ← GetBestSol(Population)
for t = 1 : generation iteration
    for k = 1 : K
        α_k^p1, α_k^p2 ← SelectParents(Population, generation iteration)
        α_k^offspr ← mutation(crossover(α_k^p1, α_k^p2))
        If (s^k(α_k^offspr))^* > (s^k(α_k))^* // EvaluatePopulation
            α_k^* = α_k^offspr // GetBestSol
    End for
    Population ← Replace(Population, α_k^*)
End for
```

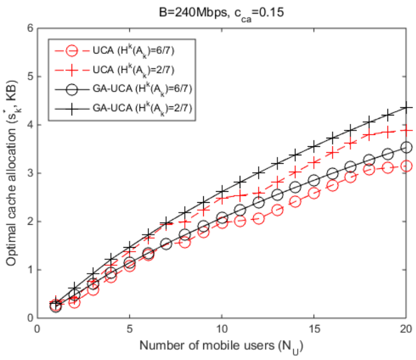
[Fig. 2] GA-UCA Algorithm

First, offspring αoffsprki inherits a part of sequences for αki from two parents using the uniform crossover. Next, using mutation one or zero value changes into zero or one value as pair dependent on Hk(Ak).

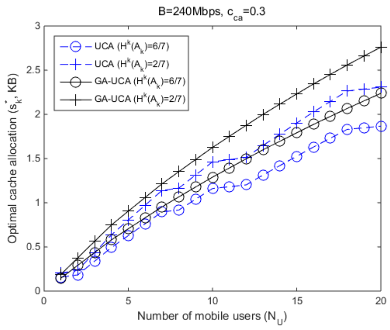
3. Numerical Results

This section presents analytical results for the optimization problem (1) using the proposed UCA and GA−UCA. We position MRs using DistMesh, which is a simple mesh generator in MATLAB[16]. Coordinates (x,y) for all mobile users are randomly generated near the MRs. Mobile users randomly decide the content to watch proportional to content popularity based on Zipf−like distribution[7]. System parameters are set as followings: NMR = 7, K = 7, η = 9, L = 8KB, and Pb = 10−3. NU gradually increases from 1 to 20.

Fig. 3 shows optimal cache allocation between UCA and GA−UCA for given high link rate (B = 240Mbps) and low cache cost (cca = 0.15). Except that NU = 1, the GA−UCA scheme assigns more



[Fig. 3] Optimal cache allocation between UCA and GA−UCA for given low cache cost



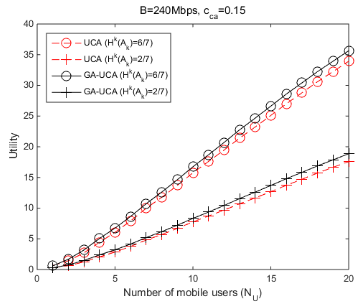
[Fig. 4] Optimal cache allocation between UCA and GA−UCA for given high cache cost

cache for each content than the UCA scheme due to reduced ETT by GA and high link rate.

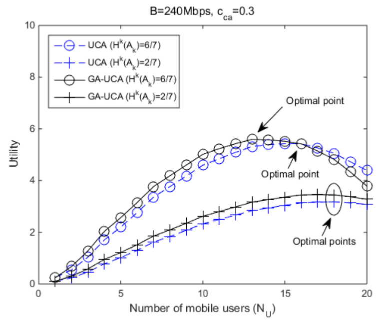
Fig. 4 shows optimal cache allocation between UCA and GA−UCA for given high link rate (B = 240Mbps) and high cache cost (cca = 0.3). Except that NU = 1, the GA−UCA scheme assigns more cache for each content than the UCA scheme. Compared with Fig. 3, due to cache cost, optimal cache is assigned slightly low. However, the gap between GA−UCA and UCA for the given cache hit ratio in Fig. 4 is slightly large due to cache cost.

Fig. 5 shows utility between UCA and GA−UCA for given high link rate (B = 240Mbps) and low cache cost (cca = 0.15). For UCA scheme, the utility is slightly higher than the GA−UCA scheme by Fig. 3. A WMN provider can accommodate more than 20 mobile users. Thus, since the utility have the greatest values,
high link rate and low cache cost can achieve the most satisfaction.

Fig. 6 shows utility between UCA and GA−UCA for given high link rate (B = 240Mbps) and high cache cost (cca = 0.3). Compared with Fig. 5, the maximum points for both schemes are shown. This means that a WMN provider can decide the number of mobile users to achieve the maximum satisfaction.



[Fig. 5] Utility between UCA and GA-UCA for given low cache cost



[Fig. 6] Utility between UCA and GA-UCA for given high cache cost

When the given cache hit ratio is high, optimal points for the number of mobile users are ranged from 13 to 16. Otherwise, optimal points for the number of mobile users reach to 18.

4. Conclusion

This paper has dealt with the optimization problem for user centric cache allocation, considering content popularity and the volume of mobile users connected to the near MRs in infrastructure WMNs. Besides, we also have presented a GA-UCA scheme to revise mobile users' ETT as well as improve the cache allocation of the UCA scheme. Finally, our proposed schemes provide the optimal cache allocation and the optimal number for mobile users.

REFERENCES

[1] I. F. Akyildiz & X. Wang. (2005). A survey on wireless mesh networks. *IEEE Communications magazine, 43(9),* S23-S30.
DOI : 10.1109/MCOM.2005.1509968

[2] Zhu, Y., Zeng, W., Liu, H., Guo, Y & Mathur, S. (2008, May). Supporting video streaming services in infrastructure wireless mesh networks: architecture and protocols. *In 2008 IEEE International Conference on Communications (pp. 1850-1855). IEEE.*
DOI : 10.1109/ICC.2008.355

[3] A. Alasaad, S. Gopalakrishnan & V. C. Leung. (2015). Replication schemes for peer-to-peer content in wireless mesh networks with infrastructure support. *Wireless Communications and Mobile Computing, 15(4),* 699-715.
DOI :10.1002/wcm.2376

[4] R. K. Sheshadri & D. Koutsonikolas. (2013). Comparison of routing metrics in 802.11 n wireless mesh networks. In 2013 Proceedings *IEEE INFOCOM (pp. 1869-1877). IEEE.*
DOI : 10.1109/INFCOM.2013.6566986

[5] S. M. Das, H. Pucha & Y. C. Hu. (2007). Mitigating the gateway bottleneck via transparent cooperative caching in wireless mesh networks. *Ad hoc networks, 5(6),* 680-703.
DOI : 10.1016/j.adhoc.2006.11.004

[6] X. Cai, S. Zhang & Y. Zhang. (2013, April). Economic analysis of cache location in mobile network. In 2013 IEEE Wireless Communications and Networking Conference *(WCNC) (pp. 1243-1248). IEEE.*
DOI : 10.1109/WCNC.2013.6554742

[7] S. H. Jeon, S. An, C. Yoon, H. W. Lee & J. Choi. (2015). User centric content management system for open IPTV over SNS. *Journal of Communications and Networks, 17(3),* 296-305.
DOI :10.1109/JCN.2015.000052

[8] S. Borst, V. Gupta & A. Walid. (2010). Distributed caching algorithms for content distribution networks. In 2010 Proceedings *IEEE INFOCOM (pp. 1-9). IEEE.*
DOI : 10.1109/INFCOM.2010.5461964

[9] Y. Wang, Z. Li, G. Tyson, S. Uhlig & G. Xie. (2015). Design and evaluation of the optimal cache allocation for content-centric networking. *IEEE Transactions on Computers, 65(1),* 95-107.
DOI :10.1109/TC.2015.2409848

[10] A. Vakali. (2002). Evolutionary techniques for web caching. *Distributed and Parallel Databases, 11(1),* 93-116.
DOI : 10.1023/A:1013385708178

[11] L. Wang. (2011). *Modeling and Designing Fair Rate Control for Wireless Mesh Networks with Partial Interference.*

[12] S. Boyd & L. Vandenberghe. (2004). *Convex optimization.* Cambridge university press.

[13] L. Fan, P. Cao, J. Almeida & A. Z. Broder. (2000). Summary cache: a scalable wide-area web cache sharing protocol. *IEEE/ACM Transactions on Networking (TON), 8(3),* 281-293.
DOI : 10.1109/90.851975

[14] R. Draves, J. Padhye & B. Zill. (2004, September). Routing in multi-radio, multi-hop wireless mesh networks. In Proceedings of the 10th annual international conference on Mobile computing and networking (pp. 114-128). ACM.
DOI :10.1145/1023720.1023732

[15] Y. Lin & V. W. Wong. (2006, November). WSN01-1: frame aggregation and optimal frame size adaptation for IEEE 802.11 n WLANs. In IEEE globecom 2006 (pp. 1-6). IEEE.
DOI :10.1109/GLOCOM.2006.925

[16] P. O. Persson & G. Strang. (2004). A simple mesh generator in MATLAB. *SIAM review, 46(2),* 329-345.
DOI :10.1137/S0036144503429121

전 승 현(Seung-Hyun Jeon) [정회원]

· 1999년 2월 : 부경대학교 전자계산학과(이학학사)
· 2009년 8월 : 한국과학기술원 정보통신공학과(공학석사)
· 2017년 2월 : 한국과학기술원 전기및전자공학부(공학박사)
· 2017년 3월 ~ 현재 : 한국과학기술원 박사후연구원
· 관심분야 : 분산망, 에너지 절감
· E-Mail : creemur@kaist.ac.kr