M

Mobile Edge Caching in HetNets

Xiuhua Li[1], Xiaofei Wang[2], and Victor C. M. Leung[1]
[1]The University of British Columbia, Vancouver, BC, Canada
[2]School of Computer Science and Technology, Tianjin University, Jinnan, Tianjin, China

Synonyms

HetNet; Mobile edge caching

Definition

Mobile edge caching refers to distributing content files (e.g., videos, audios, photos, application programs, and so on) from service providers over the Internet to caches that are deployed at the edges (e.g., mobile devices and base stations) of mobile networks, aiming at bringing content closer to mobile users in the distance of the network topology to deal with the challenge of the explosive growth in content requests from users in mobile networks. HetNet is short for heterogeneous network and is a form of radio access networks (RANs) with complex interoperation between macro cells and small cells, which consists of different types of base stations (BSs) such as femto BSs, pico BSs, micro BSs, and macro BSs.

Mobile edge caching in HetNets is the combination between the technique of mobile edge caching and the network architecture of HetNets, aiming to achieve their joint benefits in enhancing network performances. Due to the high disparity of content popularity (i.e., a small number of content files actually may attract a large amount of downloads), by deploying hierarchical collaborative caching at the edges of HetNets, the content delivery cascades can be optimized during the intermediate transmissions, while reduced content delivery delays can be also provided.

Historical Background

With the rapid advancement of mobile networks from 2G and 3G to current 4G, people's daily life has changed significantly, and people are increasingly enjoying online social activities on mobile devices (e.g., smartphones and electronic tablets). As a result, requests for various content files (e.g., videos, audios, photos, application programs, and so on) from mobile users are increasing at an explosive speed, which has become a serious issue of mobile network operators (MNOs). However, current RANs and backhaul networks cannot support these content requests effectively due to the scarcity of network resources and the limit of their network architectures (Wang et al. 2015). Thus, to deal with those issues, it is necessary to

X.(S.) Shen et al. (eds.), *Encyclopedia of Wireless Networks*,
https://doi.org/10.1007/978-3-319-32903-1_46-1

employ revolutionary schemes in network architectures and data transmissions toward the fifth-generation (i.e., 5G) mobile networks.

Mobile edge caching is regarded as an effective technique to reduce the reduplicated network traffic load in mobile networks. In particular, studies in (Cha et al. 2007; Chen et al. 2002) have shown that a large of portion of the network traffic is caused by massive duplicated downloads of the same popular content. For instance, top 10% of videos in YouTube account for about 80% of all the views (Chen et al. 2002). Thus, with this technique, by caching popular content at the edges (e.g., mobile devices and BSs) of mobile networks, the requested content can be closer to mobile users in the distance of the network topology, and mobile users can directly access the content cached at the edges instead of downloading the content from SPs over the Internet via backhaul networks. Reduplicated transmissions from servers to clients are avoided, and most of the requests from users can be satisfied intermediately in mobile networks. Consequently, mobile edge caching can effectively enhance the network performances, especially on offloading network traffic (Chen and Yang 2016; Li et al. 2016, 2017), reducing system costs (Zhi et al. 2016; Gregori et al. 2016), and improving the quality of service (QoS) or quality of experience (QoE) of mobile users (Golrezaei et al. 2012; Zhao et al. 2016; Ao and Psounis 2015; Hong and Choi 2016; Li et al. 2015).

Another effective approach is to introduce the architecture of HetNets, which consists of different types of BSs (such as femto BSs, pico BSs, micro BSs, and macro BSs) with different wireless coverages (Li et al. 2016, 2017). HetNets can greatly enhance wireless link quality between mobile users and BSs and thus improve network capacity.

Considering the great potentials of the above two techniques, it is beneficial to combine them together, i.e., mobile edge caching in HetNets, which can effectively address the above discussed issues of MNOs. Studies in (Wang et al. 2015; Chen and Yang 2016; Zhi et al. 2016; Gregori et al. 2016; Golrezaei et al. 2012; Zhao et al. 2016; Ao and Psounis 2015; Hong and Choi 2016; Li et al. 2015) focused on single-tier caching in either mobile devices or BSs in mobile networks. Studies in (Li et al. 2016, 2017; Yang et al. 2016; Jiang et al. 2017; Xu and Tao 2017) focused on hierarchical BS caching in HetNets, where the edge caching in mobile devices is not considered. Studies in (Rao et al. 2016; Wang et al. 2017) explored the mobile edge caching in both mobile devices and BSs in mobile networks.

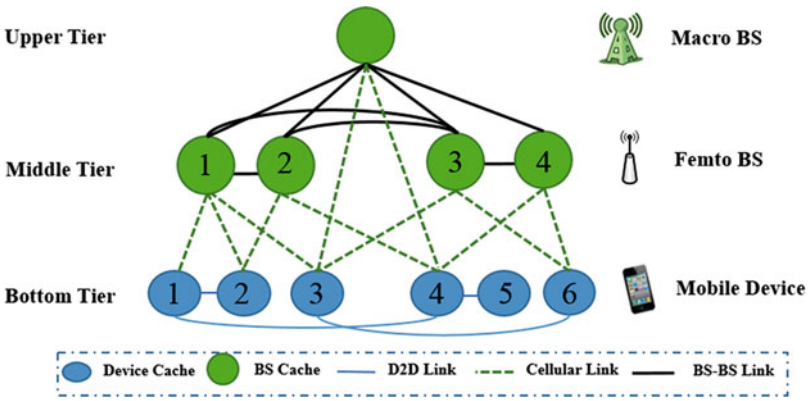
Foundations

To satisfy content requests from mobile users, mobile edge caching in HetNets needs to consider two phases, i.e., content placement phase and content delivery phase.

In the content placement phase, mobile edge caching in HetNets mainly deals with the following four issues:

- **Caching Topology** – To bring content closer to mobile users, it is important to decide where to cache in mobile networks. In HetNets, content can be cached at the edges that consist of mobile devices and BSs, which can form an either single-tier or hierarchical edge caching topology. Firstly, MNOs need to design the network architecture of HetNets, especially about how many tiers or types of BSs there are in HetNets as well as how different BSs connect. Secondly, MNOs need to consider whether mobile devices are able to cache content or not. Thirdly, MNOs need to consider which tiers of BSs are able to cache content or not. As a result, the mobile edge caching topology can be achieved, denoting the deployment of caches at the edges of HetNets.

  As illustrated in Fig. 1, the caching topology of mobile edge caching in a HetNet is hierarchical and consists of three tiers of caching, i.e., bottom tier with six mobile device caches, middle tier with four femto BS caches, and upper tier with one marco BS cache. Here, the HetNet consists of two tiers of BSs, i.e., four femto BSs and one macro BS. All the BSs and mobile devices are able to



Mobile Edge Caching in HetNets, Fig. 1 An illustration of caching topology

cache content. Besides, the connection among BSs, among mobile devices, and between BSs and mobile devices is via BS-BS links, D2D links, and cellular links, respectively. In particular, mobile devices are possible to be connected with any tier of BSs. For instance, mobile device 3 is connected with femto BS 1 and femto BS 3 at the middle tier as well as the macro BS at the upper tier.

- **Content cacheability and storage format** – Mobile edge caching in HetNets aims to achieve a trade-off between network performances (e.g., network traffic load, system costs, and QoS/QoE) that are usually expensive to be improved and storage costs that are becoming much cheaper. However, the practical scale of content owned by SPs is growing rapidly, and thus it is impossible to cache all the content. Thus, it is important to decide what content to cache (i.e., content cacheability) taking content popularity into account. As practically captured in (Cha et al. 2007; Chen et al. 2002), only a small amount of popular content accounts for a large portion of content requests from mobile users, while a long tail of content remains unpopular. Besides, different types of content have different cacheability (Wang et al. 2015). For instance, among the content types, videos and photos have the highest revisit rate. Moreover, a large content file can be chunked into a series of original segments, and original segments can be encoded into an arbitrary number of packets to explore the diversity of cached content, while a given number of different coded packets can be collected together and decoded into the original segments for recovering the content file (Golrezaei et al. 2012; Yang et al. 2016). Thus, a content file can be cached by storing in three formats: (1) entire original content in a neither unchunked nor uncoded case, (2) original segments in a chunked but uncoded case, and (3) coded packets from the original segments in a chunked and coded case.
- **Caching policy** – Caching policies, deciding what to cache, how to cache, and when to release caches for what network objectives, are crucial to achieve the performance gains of mobile edge caching. In particular, the optimization objectives of mobile edge caching in HetNets can be various for enhancing network performances, such as offloading network traffic, reducing system costs, improving mobile users' QoS/QoE, and so on. It is also important to estimate the gain of caching a content file by evaluating its current popularity, potential popularity, storage size, and locations of existing replicas in the network topology based on the system learning and analysis from mobile users' social behaviors and preferences (Li et al. 2016, 2017). Mobile edge caching policies can be operated in offline/online manners based on the practical network requirements. Offline caching (i.e., proactive) is relatively static with some prior network knowledge such as content requests and content popularity, while online

caching (i.e., reactive caching) is relatively flexible according to the dynamics of network knowledge. Rather than employing traditional caching policies (e.g., least recently used (LRU), least frequently used (LFU), and first-in first-out (FIFO)), it is important and challenging to design proper cooperative mobile edge caching policies in HetNets to improve the network performances.

- **Operation time of caching** – According to the change of content popularity and operated manners of mobile edge caching, different types of content can be cached in different time periods. Online caching needs to cache content immediately or in a short time, while offline caching does not. For instance, in offline caching, short-lifetime popular news with short videos are updated every a few hours, while long-lifetime new movies and new music videos are, respectively, posted weekly and monthly (Li et al. 2017). In order to reduce the traffic load and avoid possible network traffic congestion especially in busy hours, content can be cached in off-peak hours (e.g., late night).

In the content delivery phase, each mobile user requests content based on its own preference. In order to satisfy a user's content request, mobile edge caching in HetNets mainly deals with the following two issues:

- **Content request routing –** Content request routing shows the possible routes for delivering the requested content in the derived caching topology to a user before operating practical wireless transmissions of content in HetNets. Specifically, in the network, the whole process of content request routing can be summarized as:
  - If a user's requested content is locally cached in its mobile device, then the request can be satisfied locally.
  - Otherwise, the user can first find the content in caches of other users in close proximity and then establishes a device-to-device (D2D) link with a user where the content is available and finally fetches the content in a D2D manner (e.g., Wi-Fi Direct, or Bluetooth) (Gregori et al. 2016; Rao et al. 2016).
  - If not yet satisfied, the user has to be served by the associated BSs via cellular links. If the requested content is locally cached, then the associated BSs satisfy the request directly; otherwise, the associated BSs need to explore the cooperation possibility for fetching the content from other BSs where the content is available.
  - If not yet satisfied, then downloading the content directly from SPs over the Internet via backhaul networks is the last resort.
- **Wireless transmission of content –** After the content request routing is known, the requested content will be delivered to mobile users with wireless transmissions via either D2D links or cellular links. In terms of D2D links, they are established only when a pair of mobile users are in close proximity and willing to share the cached content in a D2D manner. Here, social behaviors and content preference of mobile users need to be analyzed and estimated in advance by system learning. In terms of cellular links, the associated BSs with noncooperation or cooperation can transmit the requested content to the user by unicasting/multicasting based on the practical data transmission schemes in HetNets. In particular, resource allocation and scheduling for wireless transmissions of content via cellular links is necessary to enhance the network performances while satisfying the content requests.

Key Applications

The technique of mobile edge caching in HetNets can be utilized for general MNOs to deal with the explosive growth in content requests from mobile users, and thus there are also venders that are manufacturing cache-enabled base station products, e.g., cache-enabled femtocell products and cache-enabled Wi-Fi routers. For companies of content delivery networks (CDNs), a new key trend is also the extension of their services into

mobile edge networks, particularly for BSs with opened interfaces for MNOs and third-party content providers. Mobile edge caching in HetNets can also be utilized in 4G networks, but will be widely deployed in 5G networks.

Future Directions

From the evolution of mobile edge caching in HetNets, it appears that, in the future, emphasis will be given on the following research directions as follows:

- Big data-based large-scale online/offline optimization of content caching for MNOs
- Machine learning-based analysis and prediction on mobile users' social behaviors and preference for caching and prefetching optimization
- More rapid and secure collaborations among mobile devices and BSs in HetNets
- Pricing methodology for mobile edge caching in HetNets

Cross-References

- ▸ Content-Centric Mobile Networks
- ▸ FemtoCaching
- ▸ Hierarchical Web Caching
- ▸ Large-scale Optimization
- ▸ Mobile Edge Computing
- ▸ Proxy Caching

References

Ao WC, Psounis K (2015) Distributed caching and small cell cooperation for fast content delivery. In: Proceedings of theACM MobiHoc, June 2015, pp 127–136

Cha M, Kwak H, Rodriguez P, Ahn YY, Moon S (2007) I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In: Proceedings of the ACM IMC, Oct 2007, pp 1–14

Chen B, Yang C (2016) Caching policy optimization for D2D communications by learning user preference. In: Proceedings of the IEEE WCNC, May 2016, pp 1–6

Chen Y, Qiu L, Chen W, Nguyen L, Katz R (2002) Clustering web content for efficient replication. In: Proceedings of the IEEE ICNP, Nov 2002, pp 165–174

Golrezaei N, Shanmugam K, Dimakis AG, Molisch AF, Caire G (2012) FemtoCaching: wireless video content delivery through distributed caching helpers. In: Proceedings of the IEEE INFOCOM, Mar 2012, pp 1107–1115

Gregori M, Vilardebó JG, Matamoros J, Gündüz D (2016) Wireless content caching for small cell and D2D networks. IEEE J Sel Areas Commun 34(5):1222–1234

Hong JP, Choi W (2016) User prefix caching for average playback delay reduction in wireless video streaming. IEEE Trans Wirel Commun 15(1):377–388

Jiang W, Feng G, Qin S (2017) Optimal cooperative content caching and delivery policy for heterogeneous cellular networks. IEEE Trans Mobile Comput 16(5):1382–1393

Li X, Wang X, Xiao S, Leung VCM (2015) Delay performance analysis of cooperative cell caching in future mobile networks. In: Proceedings of the IEEE ICC, June 2015, pp 5652–5657

Li X, Wang X, Leung VCM (2016) Weighted network traffic offloading in cache-enabled heterogeneous networks. In: Proceedings of the IEEE ICC, May 2016, pp 1–6

Li X, Wang X, Li K, Han Z, Leung VCM (2017) Collaborative multi-tier caching in heterogeneous networks: modeling, analysis, and design. IEEE Trans Wirel Commun 16(10):6926–6939

Rao J, Feng H, Yang C, Chen Z, Xia B (2016) Optimal caching placement for D2D assisted wireless caching networks. In: Proceedings of the IEEE ICC, May 2016, pp 1–6

Wang X, Li X, Leung VCM, Nasiopoulos P (2015) A framework of cooperative cell caching for the future mobile networks. In: Proceedings of the HICSS, Jan 2015, pp 5404–5413

Wang W, Lan R, Gu J, Huang A, Shan H, Zhang Z (2017) Edge caching at base stations with device-to-device offloading. IEEE Access 5:6399–6410

Xu X, Tao M (2017) Modeling, analysis, and optimization of coded caching in small-cell networks. IEEE Trans Commun 65(8):3415–3428

Yang C, Yao Y, Chen Z, Xia B (2016) Analysis on cache-enabled wireless heterogeneous networks. IEEE Trans Wirel Commun 15(1):131–145

Zhao Z, Peng M, Ding Z, Wang W, Poor HV (2016) Cluster content caching: an energy-efficient approach to improve quality of service in cloud radio access networks. IEEE J Sel Areas Commun 34(5):1207–1221

Zhi W, Zhu K, Zhang Y, Zhang L (2016) Hierarchically social-aware incentivized caching for D2D communications. In: Proceedings of the IEEE ICPDS, Dec 2016, pp 316–323