A Survey on Caching in Distributed Small Cell Networks

M. Venu Gopalachari
Assistant Professor, Department of CSE
Chaitanya Bharathi Institute of Technology
Hyderabad, India
Email:venugopal.m07@gmail.com

***Abstract*—**The exponential growth of mobile devices such as smartphones and tablets, coupled with proliferation of online social networks has considerably increased the traffic in cellular networks. In contrast to classical cellular traffic that was only based on voice and audio communications, the recent technologies have resulted in bandwidth-intensive services such as video streaming, and video conferencing increases the traffic among users. This traffic surge affects the capacity of existing wireless networks which makes it difficult to ensure the high quality-of-service (QoS) required by the cellular services. In order to handle with the limited capacity of existing cellular networks and keep up with the strict QoS requirements, in terms of data rate and delay tolerable application-specific delays, a new generation of wireless networks has emerged. To achieve the requirements of this new generation and provide efficient infrastructure support for this data deluge, several research challenges must be addressed and solved. In this paper a survey on literature about small cell networks in distributed environment is presented which focus on caching aspect to improve the performance. The related work for caching in distributed small cell networks is also presented.

Keywords- *Small cell Network; Cache Management; Distributed Cache; Network Traffic.*

*****

I. INTRODUCTION

To support the growth of mobile traffic demands, mobile network operators have evolved their networks from classical wireless networks that are based on the sparse deployment of macro base stations (MBSs) to heterogeneous small cell networks (SCNs) [1]. In SCNs, a large number of small base stations of different types, such as picocells and femtocells, is densely deployed closer to the end-users. These short range and low power small base stations are used to overload the MBSs or complement their coverage [2]. Based on the concept of frequency reuse in SCNs, a large number of small cells may reuse the same spectrum locally resulting in increased spectral efficiency, improved signal quality, better cellular coverage in rural and isolated areas, and higher capacity at the edge of the macro cells. There are three major challenges of SCN's and differences with SBN's.

First one is regarding energy consumption, in which the MBSs are continuously powered, in SCNs, it is not always possible to provide grid power to all SBSs due to their possible locations which could be hard-to-reach in rural zones, or due to the remote deployment in outdoor environments. Moreover, the introduction of millimetre wave backhaul in small cell networks induces a significant increase of power consumption compared to classical wireless networks. In fact, the SBSs require more power to perform the beamforming techniques which are necessary for transmissions over high frequency bands, especially when the SBSs are equipped with multiple antennas. Thus, in contrast to classical cellular networks in which the energy efficient techniques ignore the processing power and only account for optimizing the transmit power, new energy efficient approaches need to be developed for SCNs.

Second difference is dense deployment of SBSs One of the solutions for boosting the capacity of wireless networks, is the extensive spatial reuse through the dense orthogonal deployment of small cells. In fact, the number of SBSs is already larger than the number of deployed MBSs and is envisioned to exceed the number of mobile devices. The dense deployment of small sized cells creates more boundaries resulting in more frequent handovers to support mobile users compared to traditional cellular networks. This comes with the expense of significant signalling overhead, transmission delays and packet losses which affect considerably the QoS experienced by the end-users. Classical handover techniques that are based on the received signal power are less efficient in SCNs as all SBSs with nearly equal strength will result in very frequent handovers between the multiple small cells.

The third difference is the heterogeneity of the backhaul that is in SCNs, the SBSs cannot all be connected to the core network via fibre backhaul due to the cost and geographical deployment challenges in urban environments. The SBSs are thus connected to the core network via capacity-limited wireless backhaul links which makes it difficult for the SBSs to meet the QoS requirements of the users, especially during peak hours. The capacity of conventional wireless networks is inherently limited because of the limited licensed spectrum. Thus, to support the entire traffic surge, MNOs have started considering the use of the entire available and underused spectrum such as high frequency bands via millimetre wave communication and the exploitation of unlicensed bands. This results in a heterogeneous wireless backhaul with different characteristics making high capacity and reliable communications via the backhaul one of the key challenges of SCNs deployment.

II. DISTRIBUTED CACHING IN SMALL CELL NETWORKS

The idea of distributed caching has emerged based on two observations, the nature of the traffic popularity and the availability of storage space at very low prices. In fact, only 20% of the content is responsible of more than 80% of the traffic load while the prices of storage units with a capacity in the order of terabits keep decreasing. These two factors have led to the concept of distributed caching which consists in equipping the SBSs with storage units to reduce the amount of redundant requests that are served via the backhaul [3]. For this, the SBSs predict the requests of users and download the most popular files ahead of time during peak-o_ hours, and cache these files in their storage units. Thus, the SBSs can serve most of the requests locally without using the backhaul allowing to significantly reducing the backhaul load during peak hours. The idea of distributed caching was further extended to users' devices with the introduction of device-to-device (D2D) communications as part of future cellular networks. Such network architecture allows users that are endowed with large storage capacities to cache content in their devices and share the cached content by communicating directly with one another via D2D communications. An illustration of a caching system is provided in Figure 1.

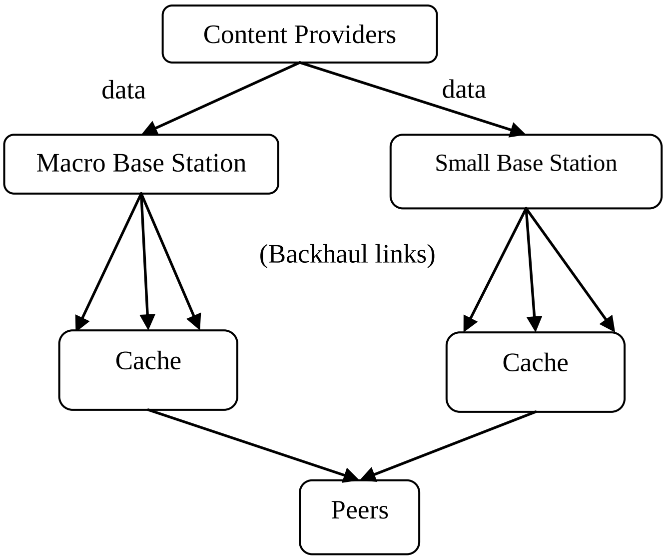

Figure 1: Cache Enabled Small Cell Network

To successfully deploy SCN caching solutions, MNOs require the cooperation of the CPs that have to share their contents with the MNOs and provide these contents' global popularity to the MNOs [4]. However, although CPs can improve the QoS of their users by caching, they might be reluctant to share their content with the MNOs. This can be due to reasons such as privacy since, once the content cached at the SBSs, the MNOs can get access to all the CP's files as well as the traffic dynamics of the users subscribed to the CP. Thus, the MNOs must provide incentives to the CPs to share their data and cache it at the SBSs by introducing suitable economic arrangements that can be beneficial for both the MNO and the CPs. On the other hand, the gain of the users from the caching process depends on the caching policy that is used by the MNO which consists in determining which file to cache at each of the SBSs and when. To develop efficient caching policies, the MNOs need to account for multiple network properties at the SBSs such as the dynamics of the content popularity, the position of the base stations (BSs), their traffic load as well as the limited capacity of their storage units. Moreover, the contents requested by user’s exhibit significant similarities that can be captured by exploiting their social interactions in online social networks (OSNs).

III. RELATED WORK

Distributed caching in cellular networks beyond caching in small cells, there has been considerable works on caching in the computer science community. The closest caching models to the considered one in this thesis, is caching in content delivery networks and content centric networks [5]. The idea consists in storing data at the closest proxy servers of the content delivery networks to the end users, known as the network edge. The goal from this approach is to balance the load over the servers, reduce the bandwidth requirements and thus reduce the users service time. Content centric networks rely on the same idea of caching with more intelligent forwarding strategies. Indeed, the content _les are identified by name instead of their location, allowing to spread the content all over the Internet network in a smart way [6]. Recently, the idea of caching was introduced in cellular networks to deal with the capacity-limited backhaul in small cell networks. Despite the similarities with caching in the Internet, the network structure of SCNs is significantly different from internet architecture. Thus, new challenges arise in SCNs such as accounting for channel characteristics and interference that make the previously proposed approaches for the Internet not applicable. This led to the recent emergence of a large literature that aims to address the caching problem while taking into account the specific characteristic of SCNs, as discussed previously. Here, we classify these works based on their similarities and directions.

A. Caching and content popularity estimation

The optimality of caching policies in SCNs relies on the accuracy of files popularity estimations. Several works have addressed this problem and showed the key role of exploiting contextual information such as users' content viewing history, social connections that can be made available from D2D interactions and online social networks. In this regard, the authors in [7] proposed a transfer-learning approach in which the SBSs can exploit the social interactions between users, referred as source domain, to estimate the popularity of the files, referred as target domain. It is shown that the proposed model engenders significant backhaul overloading gains and resource savings as compared to random caching, and estimation methods that are based on collaborative filtering. In [8], the authors proposed a coded caching framework in which the SBS learn the popularity of the files based on reinforcement learning algorithms. To determine the caching policy based on the popularity estimation, the problem is formulated as a linear program. A convex relaxation of the problem is then derived and solved. The social interactions between users are exploited in [9] where users have caching capabilities and can share the content via D2D communications. In this system, the SBSs predict the set of most influential users and cache the _les at the users and SBSs based on files' popularity. The impact of the correlation of users' interest on the caching policy is considered in [10]. The authors in [11] assumed that the SBSs do not have any information about the files popularity and can only observe the instantaneous demand. Using multi-armed bandit theory, three algorithms are proposed to define the caching policies without knowing the popularity of the files. Similarly in [12], the authors address the problem of cache placement when only the instantaneous estimation of the content popularity is known at the SBSs. A centralized algorithm is proposed to simultaneously estimate files popularity and determine the most efficient caching policy.

B. Coded Caching

Coded caching enables the SBSs to cache fractions of a given a file at multiple SBSs. To this end, the files are coded using different coding schemes such erasure generating codes or maximum-distance separable (MDS) codes. Thus, a user can be served from multiple SBSs and recover the original file after collecting a given number of fractions of the _le that it requested. The idea of coded caching was introduced in [13] where the authors showed that in contrast to single-SBS caching systems, caching the most popular files in a SCN composed of multiple SBSs is not the most optimal caching policy. A grouped coded caching scheme is proposed to minimize the expected load of the backhaul links while assuming that the popularity of the files follow a heavy-tailed distribution.

An information theoretic formulation of the caching problem is introduced in [14] and a coded caching scheme that exploits both local and global gains of caching is proposed. The fundamental lower bounds on the rate required for serving the files, using the proposed coding scheme, are derived and the rate of the user is shown to increase within a factor compared to not coded caching schemes. Two hierarchical coded caching schemes with two caching levels are proposed in [15]. The first approach provides coded multicasting opportunities within each level while the second approach provides coded multicasting opportunities across multiple levels. The work in [16] proposed an online caching scheme and proved that the performance of the optimal online scheme is comparable with the performance of the optimal offline scheme, in which the storage units are refreshed based on the entire set of popular files before each new request.

In another line of works, coded caching policies were proposed in MIMO systems to exploit the cooperation possibilities. In [17], the joint problem of power and encoded cache control is formulated using an optimization approach. The goal in this model is to create more opportunities for serving users cooperatively by multiple SBSs that are equipped with multiple antennas.

C. Caching in D2D-enabled networks

The large available storage space at the users' devices has motivated the extension of the idea of caching into users devices. Thus, a given user can cache the most popular files and serve the users in its velocity via D2D communications. The authors in [18] used tools from stochastic geometry to derive the probability of successful transmissions in cache enabled networks in the presence of D2D communications. An optimization problem is then formulated to determine the best caching policy that maximizes the probability of successful file transmissions. An experimental analysis of these systems is conducted in [19] while accounting for the mobility of devices. In another line of works, the performance of caching in SCNs with enabled D2D connections is addressed from an information theoretic point of view [20]. The authors in [21] proposed a decentralized random caching placement in SCNs where a device can be served either directly or via multi-hop D2D communications from the closest device that caches the requested file. A repair scheduling is introduced to recover the lost content when a device leaves the network.

In [22], a joint encoded caching and routing problem is formulated and then reduced to a tractable facility location problem. The authors in [23] analysed the energy efficiency in networks with a macro BS and multiple SBSs equipped with caches. While accounting for content popularity and request arrival density, it is shown that the energy efficient cannot be improved when the SBSs are equipped with large sized storage units. The work in [24] analysed the impact of caching on the energy consumption in SCNs using stochastic geometry. In [25], the authors analysed the maximal energy efficiency gain brought by caching and showed that caching a number of content that exceeds a given threshold, the SBSs may not achieve higher energy efficiency. The work in [26] proposed a caching policy that optimizes the overall energy consumption of the SBSs while accounting for the multicast opportunities. The authors in [27] exploited the interdependence between caching and the MAC layer coordination. In particular, the caching problem that is considered accounts for the cooperation of multiple SBSs that can serve the same users simultaneously. The multiplexing gain is analysed and the performance improvement in terms of delivery delay is shown. An analysis of cooperation schemes using stochastic geometry is provided in [28].

Conclusion

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression, "One of us (R.B.G.) thanks . . ." Instead, try "R.B.G. thanks". Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

References

[1] T. Q. Quek, G. de la Roche, and M. Kountouris, "Small cell networks: Deployment, PHY techniques, and resource management", Cambridge University Press, 2013.
[2] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," IEEE Wireless Communications, vol. 23, no. 1, pp. 72-79, 2016.
[3] O. Fratu, A. Vulpe, R. Craciunescu, and S. Halunga, "Small cells in cellular networks: Challenges of future hetnets", Wireless Personal Communications, vol. 78, no. 3, pp. 1613-1627, 2014.
[4] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," to appear in IEEE Journal on Selected Areas in Communications, 2016.
[5] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," IEEE Communications Magazine, vol. 52, no. 2, Pp. 131-139, 2014.
[6] A. Araldo, M. Mangili, F. Martignon, and D. Rossi, "Cost-aware caching: optimizing cache provisioning and object placement in ICN,"in IEEE Global Communications Conference (GLOBECOM), Pp. 1108-1113, 2014.
[7] E. Ba stufig, M. Bennis, and M. Debbah, \A transfer learning approach for cache-enabled wireless networks," in 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Pp. 161-166, 2015.
[8] A. Sengupta, S. Amuru, R. Tandon, R. M. Buehrer, and T. C. Clancy, "Learning distributed caching strategies in small cell networks," in 11th International Symposium on Wireless Communications Systems (ISWCS), Pp. 917-921, 2014.
[9] E. Ba_stu_g, M. Bennis, and M. Debbah, "Social and spatial proactive caching for mobile data overloading," in IEEE international conference on communications workshops (ICC), Pp. 581-586,2014.
[10] Z. Chen and M. Kountouris, "Cache-enabled small cell networks with local user interest correlation", in IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Pp. 680-684, 2015.
[11] P. Blasco and D. G□und□uz, "Learning-based optimization of cache content in a small cell base station," in IEEE International Conference on Communications (ICC), Pp. 1897-1903, 2014.

[12] B. B. Nagaraja and K. G. Nagananda, “Caching with unknown popularity profiles in small cell networks," in IEEE Global Communications Conference (GLOBECOM), Pp. 1-6, 2015.

[13] U. Niesen and M. A. Maddah-Ali, “Coded caching with nonuniform demands," in IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Pp. 221-226, 2014.

[14] M. A. Maddah-Ali and U. Niesen, “Fundamental limits of caching," IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2856-2867, 2014.

[15] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. N. Diggavi, “Hierarchical coded caching," IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3212-3229, 2016.

[16] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen, “Online coded caching", IEEE/ACM Transactions on Networking, vol. 24, no. 2, pp. 836-845, 2016.

[17] A. Liu and V. K. Lau, “Cache-enabled opportunistic cooperative MIMO for video streaming in wireless systems," IEEE Transactions on Signal Processing, vol. 62, no. 2, pp. 390-402, 2014.

[18] D. Malak and M. Al-Shalash, “Optimal caching for device-to-device content distribution in 5G networks," in 2014 IEEE Globecom Workshops (GC Wkshps), 2014, pp. 863-868.

[19] V. A. Siris and D. Dimopoulos, “Multi-source mobile video streaming with proactive caching and D2D communication," in IEEE 16th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Pp. 1-6, 2015.

[20] M. Ji, G. Caire, and A. F. Molisch, “Wireless device-to-device caching networks: Basic principles and system performance," IEEE Journal on Selected Areas in Communications, vol. 34, no. 1, pp. 176-189, 2016.

[21] S.-W. Jeon, S.-N. Hong, M. Ji, and G. Caire, “Caching in wireless multihop device-to-device networks”, in 2015 IEEE International Conference on Communications (ICC), Pp. 6732-6737, 2015.

[22] K. Poularakis, G. Iosi_dis, and L. Tassiulas, “Approximation algorithms for mobile data caching in small cell networks," IEEE Transactions on Communications, vol. 62, no. 10, pp. 3665-3677, 2014.

[23] D. Liu and C. Yang, “Will caching at base station improve energy e_ciency of downlink transmission?" in IEEE Global Conference on Signal and Information Processing (GlobalSIP), Pp.173-177, 2014.

[24] B. Perabathini, E. Bafistufig, M. Kountouris, M. Debbah, and A. Conte, “Caching at the edge: a green perspective for 5G networks," in 2015 IEEE International Conference on Communication Workshop (ICCW), Pp. 2830-2835, 2015.

[25] D. Liu and C. Yang, “Energy efficiency of downlink networks with caching at base stations," IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 907-922, 2016.

[26] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, “Exploiting caching and multicast for 5G wireless networks", IEEE Transactions on Wireless Communications, vol. 15, no. 4, pp. 2995-3007, 2016.

[27] W. C. Ao and K. Psounis, “Distributed caching and small cell cooperation for fast content delivery", in 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Pp.127-136, 2015.

[28] S. H. Chae, J. Y. Ryu, T. Q. Quek, and W. Choi, “Cooperative transmission via caching helpers”, in IEEE Global Communications Conference (GLOBECOM), Pp. 1-6, 2015.