Fog Radio Access Networks: Mobility Management, Interference Mitigation, and Resource Optimization

Haijun Zhang, Yu Qiu, Xiaoli Chu, Keping Long, and Victor C.M. Leung

Abstract

In order to make Internet connections ubiquitous and autonomous in our daily lives, maximizing the utilization of radio resources and social information is one of the major research topics in future mobile communication technologies. FRAN is regarded as a promising paradigm for the fifth generation of mobile networks. FRAN integrates fog computing with RAN and makes full use of the edge of networks. FRAN would be different in networking, computing, storage, and control compared to conventional RAN and the emerging cloud RAN. In this article, we provide a description of the FRAN architecture, and discuss how the distinctive characteristics of FRAN make it possible to efficiently alleviate the burden on the fronthaul, backhaul, and backbone networks, as well as reduce content delivery latencies. We focus on the mobility management, interference mitigation, and resource optimization in FRAN. Our simulation results show that the proposed FRAN architecture and the associated mobility and resource management mechanisms can reduce the signaling cost and increase the net utility for the RAN.

Introduction

Nowadays, billions of mobile users receive seamless and stable wireless services supported by the communication infrastructures. With the rapid development of mobile communications, dozens of network standards have emerged, including the Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) standards. In addition to advanced multiple-input multiple-output (MIMO) technologies [1], small cells, and heterogeneous networks [2], cloud radio access networks (CRANs) have emerged as a popular technology for future mobile networks [3]. CRAN features centralized resource management, with all the computing, control, and data storage of the network gathered into the cloud. The centralized data centers, cellular core networks, and backbone networks are equipped with computing, storage, and network management functions. However, recent research [4, 5] has shown that the completely centralized architecture of CRAN makes it hard to cope with the unpredictable mobility of users, the increasing density of base stations (BSs) [6], and the explosive growth of user data demand. The planning and optimization of heterogeneous networks are facing complicated inter-cell interference problems and increased management complexities. More recently, the heterogeneous cloud radio access network (HCRAN) has been proposed [7], where remote radio heads (RRHs) working in coordination with high-power nodes can effectively mitigate co-channel interference. Although HCRAN may offer better cost efficiency than CRAN [7], the complex cost structure behind HCRAN, how its resource optimization should be supported by the baseband unit (BBU) pool, and its traffic burden on the cloud center require more in-depth studies. Since all information is exchanged through the BBU pool, HCRAN may cause additional burden on the fronthaul and backhaul links, especially the wireless ones [8], compared to CRAN. In the meantime, more data is generated from various social media platforms due to their increasing popularity. Hence, it becomes increasingly important to consider social networking and local information in the management and optimization of RANs. This is not easily achieved in CRAN or HCRAN because of their centralized architecture.

In view of the above issues related to CRAN and HCRAN and the requirements of the future communication scenarios, fog RAN (FRAN) was introduced by Cisco to exploit local signal processing and computing, cooperative resource management, and distributed storing/caching capabilities at the network edge [9]. In FRAN, a large amount of signal processing and computing is performed in a distributed manner, rather than all by the centralized BBU pool, and local data can be stored in edge devices, such as access points (APs) and user equipment (UE), instead of the cloud data center. A unique feature of FRAN is to maximize the use of edge devices of the network, for example, to perform collaborated radio signal processing. As a result, the burden on the fronthaul is relieved more than in CRAN or HCRAN. Due to these distinctive characteristics of FRAN, network management and optimization mechanisms need to be revis-

Haijun Zhang and Keping Long (corresponding author) are with the University of Science and Technology Beijing; Yu Qiu is with Beijing University of Chemical Technology; Xiaoli Chu is with the University of Sheffield; Victor C. M. Leung is with the University of British Columbia.

Digital Object Identifier: 10.1109/MWC.2017.1700007

ited for FRAN. In [10], cooperative interference mitigation and handover management were studied for heterogeneous cloud small cell networks. An information-centric wireless network virtualization framework was studied in [11]. These requirements translate into a tremendous demand for bandwidth and energy consumption. MIMO is a promising solution for solving these issues as it provides extra degrees of freedom in the spatial domain, which promote a trade-off between diversity gain and multiplexing gain. Over the past few years, an enormous amount of research has been concentrated on MIMO communication [12]. However, the modest computational capabilities of mobile devices limit the MIMO gains that can be achieved in practice. An attractive alternative for realizing the performance gains offered by multiple antennas is multiuser MIMO, where a multiple-antenna transmitter serves multiple single-antenna receivers simultaneously. In fact, the combination of multiuser MIMO and distributed antennas is widely recognized as a viable technology for extending service coverage and mitigating interference [13]. Specifically, distributed antennas introduce additional capabilities for combating both path shadowing and loss by shortening the distances between the the receivers and transmitters. Furthermore, even with these powerful MIMO techniques, spectrum scarcity is still a main obstacle in providing high-speed uplink and downlink communications. In this article, we only consider the signal antenna system and consider MIMO beamforming-based interference management in our future work. However, mobility management, interference mitigation, and resource optimization for FRAN have not been sufficiently studied in the existing works.

In this article, we first present a network architecture of FRAN. Then we propose the handover management and handover procedures, which make use of edge caching, in FRAN. For effective interference and resource management, we introduce an interference-aware price and the price of using the fronthaul for caching in the network edge. The subchannel and power allocation problem is then modeled as a non-cooperative game to optimize the resource allocation in FRAN. Moreover, we analyze the signaling overhead in FRAN using numerical results. We show that the proposed interference and resource management mechanisms not only reduce the interference in the FRAN, but also enhance the net utility of the network.

System Architecture

Figure 1 shows the FRAN architecture, where the marco remote radio heads (MRRHs), small RRHs (SRRHs), and fog computing access points (F-APs) all connect to the BBU pool. We denote the smart user equipment as F-UE in the FRAN. The MRRHs are connected to the BBU pool by the backhaul links. The SRRHs and F-APs are connected to the BBU pool by the fronthaul links. The F-APs, which are unique to FRANs, integrate not only the fronthaul radio frequency but also the physical-layer signal processing functionalities and procedures of the upper layers. Thus, F-APs can implement collaborative radio signal processing locally using their adequate computing capabilities and can manage their caching memories flexibly. It is worth noting that although both SRRHs and F-APs are equipped with caching capabilities, the contents stored in them are fundamentally different: the contents cached in F-APs are highly locally popular or relevant, but not those cached in SRRHs. With the increasing popularity of location-based mobile applications, a lot of superfluous information may be generated, adding to the surging data traffic over the fronthaul between the SRRHs and the centralized BBU pool, which pushes the fronthaul links to their capacity limits. In fact, the social application data exchanged between neighboring F-UEs shows a high degree of conformity. Some social applications would only generate data traffic between F-UEs in close physical proximity. Besides, users from the same social group or having the same social interest may request the same contents over the downlink. In these cases, F-APs can provide the requested services locally by caching the popular contents. Consequently, users do not need to connect to the BBU pool every time they require data or contents.

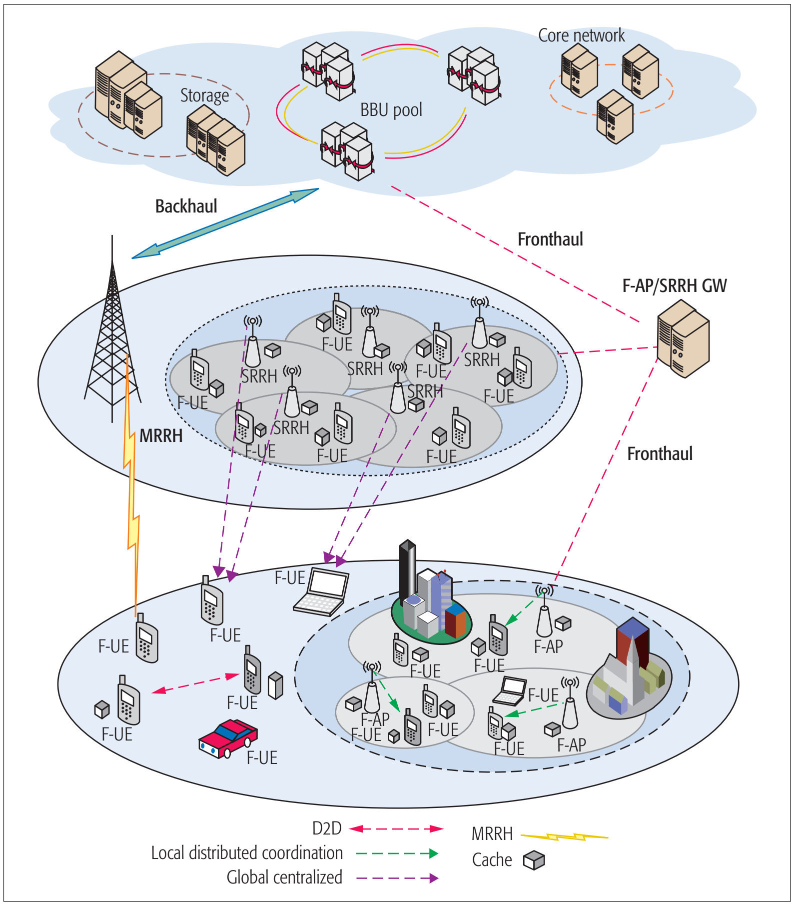

FIGURE 1. FRAN architecture.

High-mobility F-UEs are served by MRRHs. Although the interface between the BBU pool and the MRRHs on the backhaul link is compatible with that defined in the 3GPP standards for LTE and LTE-Advanced systems [14], the BBU pool will mainly provide centralized storage and communications in FRAN. SRRHs connect to the BBU pool in order to retrieve the packets cached at the BBU pool or packets from the cloud network. SRRHs are also equipped with caching capabilities to pro-

Frequent handovers also cause a heavy burden on the fronthaul, the backhaul and the core network. In order to avoid frequent handovers and alleviate the control overhead of handovers, MRRHs, SRRHs and F-APs can shift some handover related control and decision making to F-UEs in FRAN.

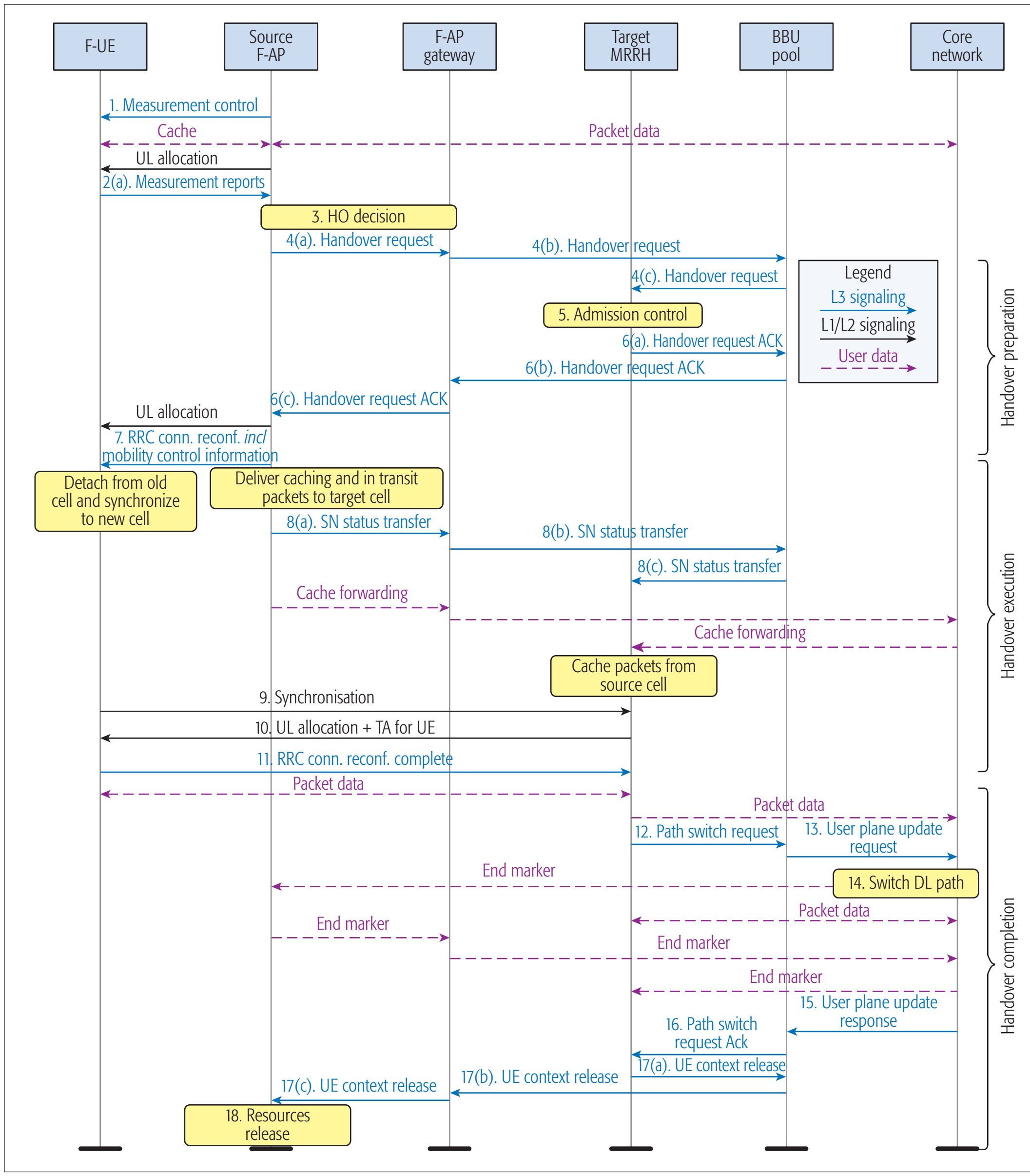

FIGURE 2. Handover from F-AP to MRRH.

vide local caching services to their associated F-UE. Collaborative signal processing can be performed by SRRHs, F-APs, and F-UEs locally, and cooperative resource management is performed by F-APs and F-UEs in a distributed manner. Consequently, the communications burden on the fronthaul and the processing and control burden on the BBU pool can be greatly relieved. The edge devices (including SRRHs, F-APs, and F-UEs) of the network are assigned some new functionalities, for instance, distributed storage, control, configuration, measurement, and resource management.

As shown in Fig. 1, there are four possible transmission modes in FRAN: device-to-device (D2D) and relay mode, local distributed coordination mode, global centralized mode, and MRRH mode. F-UEs will select the most appropriate mode through user-centric adaptive techniques, which take into account the F-UEs' movement speed, communication distance, location, quality of service (QoS) requirements, computing capability, and caching capability.

Mobility Management in FRAN

Handover Management

In mobile communications, handover management is one of the most critical techniques to guarantee the QoS requirements of users. However, handover management in FRAN has not been sufficiently studied in the existing literature. High-speed F-UEs should be served by MRRHs with large coverage areas and reliable connections. Low-mobility F-UEs should be served by SRRHs or F-APs that can provide very high capacity to a small number of F-UEs. In a heterogeneous network, unnecessary handovers (e.g., ping-pong handovers) or handover-caused radio link failures are more likely to happen compared to in conventional cellular networks due to the small coverage areas of small cells and severe co-channel interference. Moreover, frequent handovers also cause a heavy burden on the fronthaul, the backhaul, and the core network. In order to avoid frequent handovers and alleviate the control overhead of handovers, MRRHs, SRRHs, and F-APs can shift some handover-related control and decision making to F-UEs in FRAN.

Handover Procedures in FRAN

Since handover management in FRANs is different from that in heterogeneous networks or CRANs, the handover procedure should be revised for both F-AP/SRRH-to-F-AP/SRRH handover and F-AP/SRRH-to-MRRH handover. Different from the handover procedure in CRAN, many handover related functions such as handover and admission control are shifted from the BBU pool to F-APs, SRRHs, and MRRHs in FRAN, in conjunction with AP selection mechanisms for various F-UEs. Moreover, in FRAN, the data traffic is generated not only in the BBU pool but also in the network edge, such as F-UEs, F-APs, and SRRHs.

Conventional handover schemes are mainly based on the received signal strength, where handover decisions are made by comparing the received reference signal strength to a predefined threshold. This would cause many unnecessary handovers such as ping-pong handovers in FRANs. If APs of the mobile network are regarded as another kind of resource for mobile devices, the handover process can be transformed into a dynamic resource allocation problem. Note that APs may become precious or limited resources in certain scenarios, for example, mobile devices with urgent communication needs after an accident. It is reasonable to prioritize the allocation of APs to such mobile devices to guarantee the QoS of the urgent communications.

Figure 2 shows the handover procedure and signaling flow for the handover of F-UE from an F-AP to an MRRH in FRAN. This is mainly based on the handover management in heterogeneous cloud small cell networks [10], but different in that the handover decisions are made between the source F-AP and the F-AP gateway (step 3, Fig. 2) rather than in the BBU pool; and the transmission of data locally cached in an F-AP to F-UE does not need to go through the core network. The source F-AP transmits handover request signaling to target the MRRH through the F-AP gateway and BBU pool (step 4, Fig. 2). Admission control happens in the target MRRH, and then handover request Ack signaling would be transmitted to the source F-AP (step 5, Fig. 2). What follows is the data transmission between the source F-AP and the target MRRH, as shown in Fig. 2. The handover procedure in Fig. 2 can be applied to the handover from an SRRH to an MRRH as well. The procedure of handover from an MRRH to an F-AP would be the most complex, because there could be hundreds of possible target F-APs around, and there is no direct communication about the handover from the MRRH to the F-AP.

For handover between two F-APs (or two SRRHs), the handover decisions are made between the source F-AP (SRRH) and the F-AP (SRRH) gateway. The two F-APs (SRRHs) can communicate with each other through the S1 interface. The handover signaling flow between the source F-AP and the target F-AP via the F-AP gateway is shown in Fig. 3, where the FRAN architecture deploys the traditional S1 and X2 interfaces. The procedure (signaling flow) of handover between two SRRHs is similar to that in Fig. 3. Two scenarios are considered in handover between the source F-AP and the target F-AP/MRRH.

- Scenario 1: An F-UE in active state moves across the F-AP, whose session initializes out of the F-AP, and finally moves out of the F-AP.
- Scenario 2: An F-UE who initializes a session under the coverage of the source F-AP remains in the active state before moving out of the source F-AP.

The probability of scenario 1 (P_{s1}) plus the probability of scenario 2 (P_{s2}) equals the probability of the handover happening on the border of the F-AP/MRRH. The handover signaling overhead can be divided into processing overhead and transmitting overhead. The processing overhead consists of that at the F-UE, F-AP/MRRH, F-AP gateway, and core network. The transmitting overhead includes that between the F-AP and the F-AP gateway, the MRRH and the BBU pools, and the F-AP gateway and the core network. Signaling overhead of the F-AP related handover in each scenario is the product of the probability of the handover in the scenario and the signaling overhead in the scenario. The signaling overhead is assumed to be proportional to the delay required to send or process a signaling message and has no unit.

We perform computer simulation to compare the performance of the proposed FRAN handover procedure with that in a conventional RAN (denoted as non-FRAN) in terms of system signaling overhead. In the simulation, the session holding times are generated as independent random variables following an identical exponential distribution. The session arrivals follow a Poission process with average arrival rate λ. The signaling overhead is assumed to be proportional to the delay required to send or process a signaling message and has no unit. Figure 4 shows the signaling overhead vs. the average session arrival rate λ. We can see that the signaling overhead in both FRAN and non-FRAN increases with the average session arrival rate for both handover between F-APs and handover from an F-AP to an MRRH. This is because the number of handovers increases with the average session arrival rate in all considered scenarios. The non-FRAN handover procedure does not distinguish between high-speed F-UEs and low-speed F-UEs. The proposed FRAN handover procedure prevents the high-speed F-UEs from handing over from MRRHs to F-APs/SRRHs, thus avoiding unnecessary handovers. Figure 4 also shows that for either F-AP-to-F-AP or F-AP-to-MRRH handover, the handover-caused signaling overhead in FRAN is much lower than that in non-FRAN. FRAN takes full advantage of edge devices to avoid transmitting all of the data to the BBU pool and process radio signals at the SRRHs, F-APs, and F-UEs. At the same time, the handover decisions occur between the F-AP and the F-AP gateway rather than in the BBU pool; As a result, there is a significant reduction in the transmitting overhead. Also, the processing overhead in F-UE and F-AP is much smaller than that in the mobility management entity (MME) and core network. Thus, the handover procedure in FRAN leads to a significant reduction in signaling overhead and data traffic compared to the conventional RAN

Conventional handover schemes are mainly based on the received signal strength, where handover decisions are made by comparing the received reference signal strength with a predefined threshold. This would cause many unnecessary handovers such as ping-pong handovers in FRANs.

Different from conventional RANs, CRANs and HCRANs, the F-APs and F-UEs in FRANs are expected to be enhanced RRHs and UEs, respectively. F-APs integrate not only the fronthaul radio frequency, but also local collaborative radio signal processing and cooperative radio resource management capabilities.

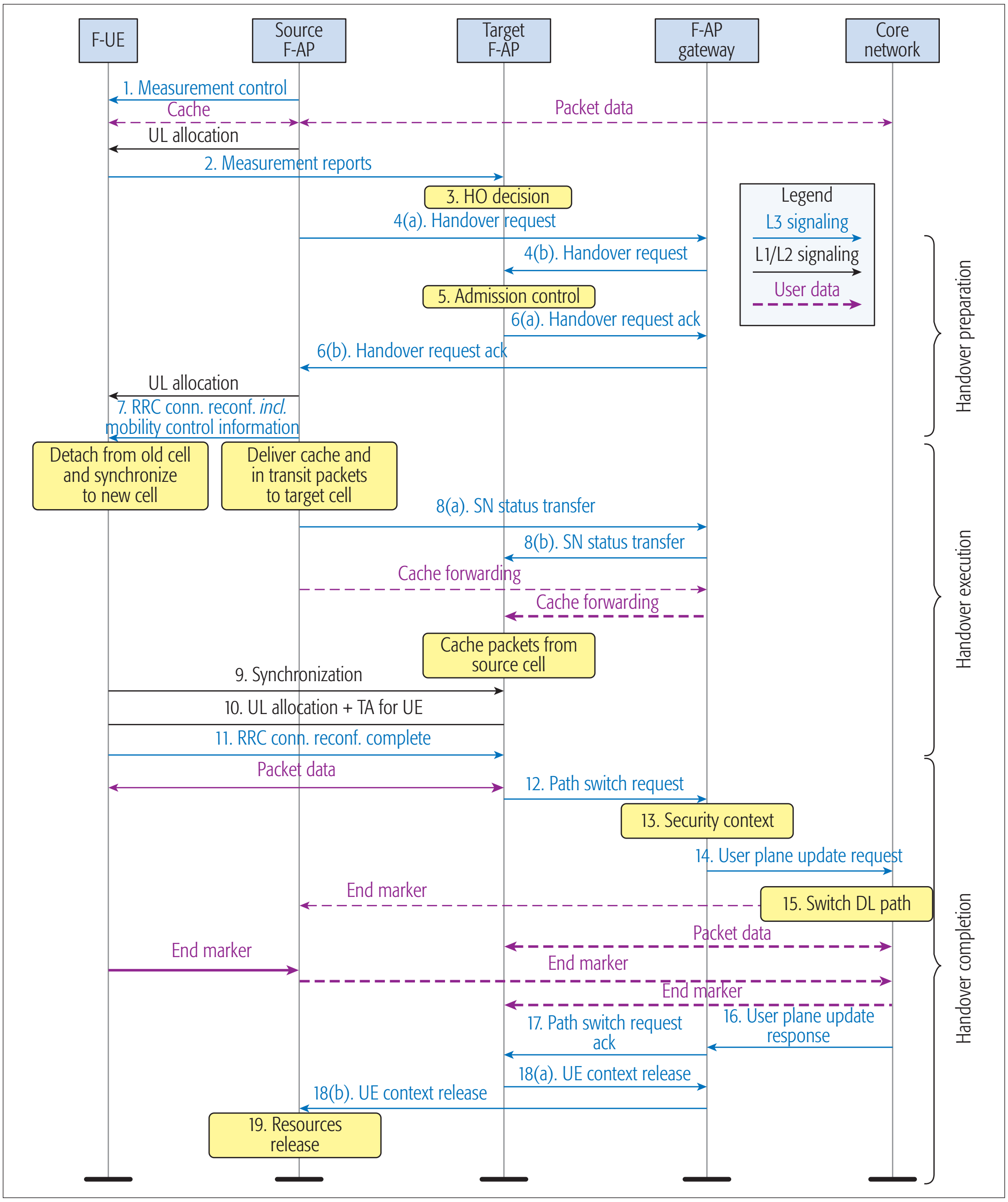

FIGURE 3. Handover from F-AP to F-AP.

and CRAN. And the long transmission delay and heavy burden on the fronthaul (usually seen in CRAN) can be alleviated effectively.

Figure 5 shows the signaling overhead vs. the average holding time for λ = 0.1. As the average session holding time increases, the total signaling overhead increases. This because the cell-boundary crossing probability increases with the session holding times, leading to a higher probability of handover.

As shown in Figs. 4 and 5, the signaling overhead for handover in FRAN is lower than that in non-FRAN. Moreover, as the MRRHs, SRRHs, and F-APs in FRAN have control functions and caching capabilities, a large portion of control signaling and data does not need to be transmitted by the BBU pool. Thus, the handover procedure in FRAN leads to a significant reduction in signaling overhead and data traffic from that in conventional RAN and CRAN.

Interference Management and Resource Optimization in FRAN

Different from conventional RANs, CRANs, and HCRANs, the F-APs and F-UEs in FRANs are expected to be enhanced RRHs and UEs, respectively. F-APs integrate not only the fronthaul radio frequency, but also local collaborative radio signal processing and cooperative radio resource management capabilities. Besides, both F-APs and F-UEs are equipped with some caching capabilities. These distinctive characteristics of FRANs ask for a revisit of resource management mechanism for FRANs. In an FRAN, F-UE can be collectively served by all the nearby SRRHs and F-APs. This means that FRAN has evolved from a BS-centric architecture to a user-centric architecture, as well as from connection-centric to content-centric. Therefore, the goal of resource optimization in FRANs is to maximize the overall communica-

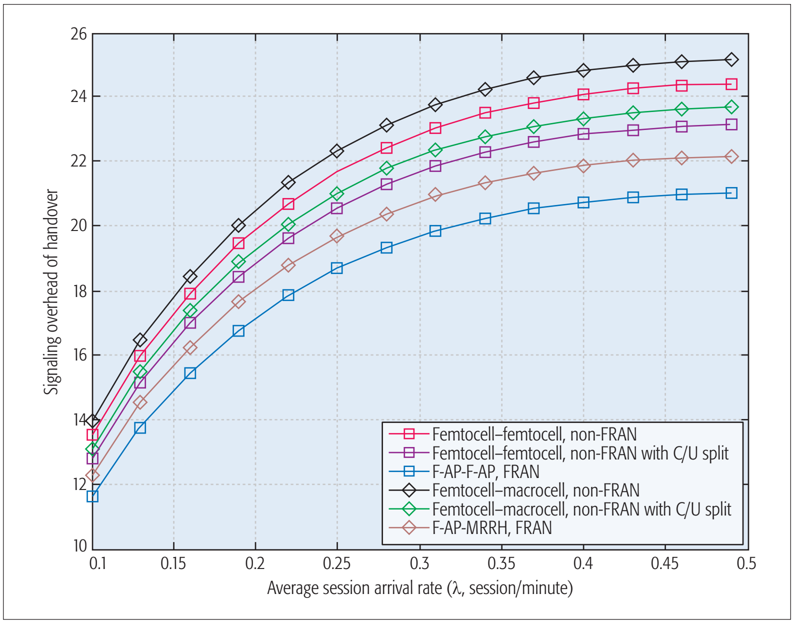

FIGURE 4. Signaling overhead vs. average session arrival rate.

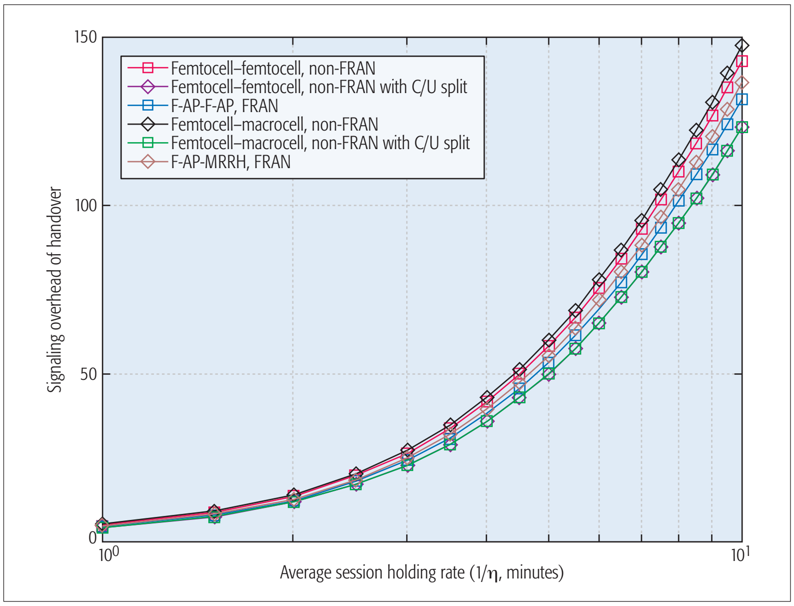

FIGURE 5. Signaling overhead vs, average session holding rate.

tion-plus-computing energy efficiency, while guaranteeing the QoS requirements on transmission rates, delays, and jitters.

In FRANs, resources are not just limited to radio resources, but also include F-APs, SRRHs, and caching and computing capabilities in F-UEs, F-APs, and SRRHs. Accordingly, the resource management in FRANs goes beyond the traditional resource allocation to also include the allocation of caching and computing resources at the network edge. As a promising approach to offload traffic from the BBU pool, the D2D and relay communications enable direct communications between mobile devices. In the following, we focus on F-APs in the discussion of resource allocation in FRANs, while the discussion can easily be extended to SRRHs.

F-APs operate either on the same frequency band as the MRRH or on a dedicated frequency band. On one hand, the dedicated channel deployment of F-APs would be difficult (if not impossible), especially when there are a large number of densely distributed F-APs. On the other hand, cross-tier interference is serious in co-channel deployment where F-APs and MRRHs share the same frequency band. Without effective management of cross-tier interference in the co-channel deployment of FRAN, both the system throughput and the energy efficiency would be largely limited. Consider a simple scenario where the FRAN contains only one MRRH and several F-APs within the coverage area of the MRRH. The F-APs share the same frequency band with the MRRH. Each F-AP makes decisions on the subchannel allocation and the power allocation on each subchannel for the F-UEs associated with it.

In previous studies, game theory has been widely applied to alleviate co-channel interference in RANs. Under the game theoretic framework, the utility function of each F-UE can be defined as the capacity (maximum achievable data rate) of the F-UE. Accordingly, the optimization of resource management in an FRAN can be formulated as the maximization of the overall network capacity under the constraint of maximal transmission power of F-UE. In this article, we model the optimization of uplink subchannel and power allocation in an FRAN as a non-cooperative game considering the selfish and rational behavior of F-UEs and F-APs. Based on the noncooperative game, we propose an interference-aware resource (power and subchannel) allocation scheme for the uplink of co-channel deployed F-APs. Particularly, we introduce a convex pricing function that is proportional to the transmission power of each F-UE in order to mitigate the interference caused by the F-UEs to the MRRH.

The proposed interference-aware uplink resource allocation scheme starts with subchannel allocation. The interference-aware subchannel allocation is achieved by maximizing the net utility function of each F-UE, which is defined as the maximum achievable data rate of the F-UE subtracted by the pricing function of the uplink interference caused by the F-UE and added by the reward function [11] of the caching offered by the F-UE. Given the optimized subchannel allocation, the power control problem is modeled as a super-modular game. It has been proven that the Nash equilibriun (NE) exists on each individual subchannel [15]. When the F-UE served by an F-AP transmits at the maximum power to achieve its maximal utility, it causes serious uplink interference to the MRRH, thus making the strategy far from Pareto-optimal. To approach the NE, an iterative scheme, in which the transmission power of each F-UE is initialized at the smallest available power level and is iteratively and sequentially updated, can be used.

We perform simulation to evaluate the performance of the proposed interference-aware uplink resource allocation scheme. In the simulation, the spectrum sharing F-APs and the F-UEs served by the MRRH are uniformly distributed in the central MRRH coverage area, and the F-UEs served by F-APs are uniformly distributed in the coverage area of their serving F-AP. The channel model of each subchannel consists of path loss and Rayleigh flat fading. In Fig. 6, we compare the total net utility function of an FRAN and that of a non-FRAN vs. the number of F-UEs (femto users) per F-AP (femtocell) for three different numbers of F-APs (femto-

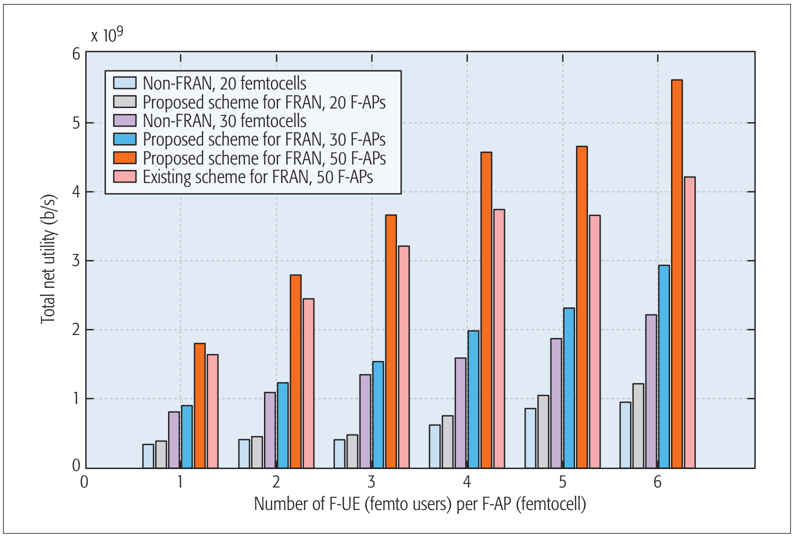

FIGURE 6. Total net utility of FRAN.

cells) in the network. The net utility gain of FRAN compared to non-FRAN performs better and better with the increasing number of F-UEs per F-AP because of the multi-user diversity. From Fig. 6, the performance of FRAN is still good with the increasing number of F-APs. In other words, the scheme we propose for FRAN can provide good services even in densely deployed networks. As shown in Fig. 6, the proposed interference-aware resource allocation scheme for FRAN performs better than the existing resource allocation for FRAN ("existing scheme for FRAN," Fig. 6) and the conventional resource allocation for LTE-A ("non-FRAN," Fig. 6). This is because in our proposed scheme based on the F-AP non-cooperative game, the interference pricing function in the net utility function can effectively mitigate the uplink interference from F-UEs to the MRRH, while the caching reward function encourages F-UEs to provide local caching, thus relieving the fronthaul between F-APs and the BBU pool and improving the total capacity of the FRAN. The total net utility of the FRAN generally increases with the number of F-UEs per F-AP, and increases with the number of F-APs.

Conclusion and Future Work

In this article, we have introduced an FRAN architecture for 5G networks. The FRAN architecture provides users with distributed local caching, computing, collaborative cooperative radio signal processing, and cooperative resource management at the edge of the network. We have proposed mobility management (including handover signaling procedures), interference mitigation, and uplink resource optimization mechanisms for FRANs. Simulation results have demonstrated that the proposed FRAN architecture in conjunction with the mobility management scheme can significantly decrease the signaling overhead of handover compared to conventional RANs. The proposed uplink resource allocation scheme based on an F-AP noncooperative game can effectively mitigate the cross-tier interference and increase the total net utility of the FRAN.

What we have discussed in this article is the foundation for FRAN. There are still many challenges and open issues that remain to be discussed in further work. For instance, software-defined networking and network functions virtualization technologies are the most effective techniques in currently practical applications. How we complete the combination of FRAN and these mature technologies is the focus of continuing study. At the same time, due to the distributed architecture of FRAN, there are some security threats in CRAN, which is made up of centralized systems. Maybe we should make the nodes in FRAN carry out selected security functions for F-UEs. The fog architectures should allow computing, storage, and networking tasks to be dynamically relocated among the fog, the cloud, and things. Therefore, the interfaces for fog to interact with the cloud, other fogs, and things and users must facilitate flexible, and in some cases dynamic, relocation of the computing, storage, and control functions among these different entities, and allow efficient and effective life cycle management of the system and services. Fog will also provide new opportunities for us to design end-to-end systems to achieve better tradeoffs between distributed and centralized architectures, between careful deployment planning and resilience through redundancy, and between what stays local and what goes global.

Now is just the beginning; we can look forward to the changes the fog will bring to the world of networking and computing in the next decades.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61471025, 61471044, 61370191), the Young Elite Scientist Sponsorship Program by CAST (2016QNRC001), and the Chinese Fundamental Research Funds for the Central Universities.

References

[1] M. Dohler *et al.*, "Is the PHY Layer Dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, Apr. 2011, pp. 159–65.
[2] A. Damnjanovic *et al.*, "A Survey on 3GPP Heterogeneous Networks," *IEEE Wireless Commun.*, vol. 18, no. 3, June 2011, pp. 10–21.
[3] Y. Lin *et al.*, "Wireless Network Cloud: Architecture and System Requirements," *IBM J. Res. Develop.*, vol. 54, no. 1, Jan. 2010, pp. 4:1–12.
[4] A. Davydov *et al.*, "Evaluation of Joint Transmission CoMP in CRAN Based LTE-A HetNets with lArge Coordination Areas," *Proc. IEEE GLOBECOM Wksps.*, Dec. 2013, pp. 801–06.
[5] M. Peng *et al.*, "Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies," *IEEE Wireless Commun.*, vol. 21, no. 6, Dec. 2014, pp. 126–35.
[6] N. Bhushanetal, "Network Densification: The Dominant Theme for Wireless Evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 82–89.
[7] S. Hung, H. Hsu, and K. Chen, "Architecture Harmonization between Cloud Radio Access Networks and Fog Networks," *IEEE Access*, vol. 3, Dec. 2015, pp. 3019–34.
[8] D. C. Chen, T. Q. S. Quek, and M. Kountouris, "Wireless Backhaul in Small Cell Networks: Modelling and Analysis," *IEEE VTC-Spring*, May 2014.
[9] F. Bonomi *et al.*, "Fog Computing and Its Role in the Internet of Things," *ACM SIGCOMM*, 2012, pp. 13–16.
[10] H. Zhang *et al.*, "Cooperative Interference Mitigation and Handover Management For Heterogeneous Cloud Small Cell Networks," *IEEE Trans. Commun.*, vol. 22, no. 3, June 2015, pp. 92–99.
[11] C. Liang *et al.*, "Virtual Resource Allocation in Information-Centric Wireless Networks with Virtualization," *IEEE Trans. Vehic. Tech.*, no. 3, Feb. 2016.
[12] F. Zhuang and V. Lau, "Backhaul Limited Asymmetric Cooperation for MIMO Cellular Networks via Semidefinite Relaxation," *IEEE Trans. Signal Processing*, vol. 62, Feb. 2014, pp. 684–93.

[13] R. Heath *et al.*, "A Current Perspective on Distributed Antenna Systems for the Downlink of Cellular Systems," *IEEE Commun. Mag.*, vol. 51, Apr. 2013, pp. 161–67.
[14] M. Peng *et al.*, "Fog-Computing-Based Radio Access Networks: Issues and Challenges," *IEEE Network*, vol. 30, no. 4, July 2016.
[15] H. Zhang *et al.*, "Interference-Aware Resource Allocation in Co-Channel Deployment of OFDMA Femtocells," *Proc. IEEE ICC*, June 2012.

Biographies

Haijun Zhang [M'13, SM'17] (haijunzhang@ieee.org) is currently a full professor at the University of Science and Technology Beijing (USTB), China. He was a postdoctoral research fellow in the Department of Electrical and Computer Engineering, University of British Columbia (UBC), Vancouver Campus, Canada. He has published more than 100 papers and authored 2 books. He serves as an Editor of *IEEE Transactions on Communications*, *IEEE 5G Tech Focus*, the *Journal of Network and Computer Applications*, and *KSII Transactions on Internet and Information Systems*, and serves/has served as a leading Guest Editor for *IEEE Communications Magazine* and *IEEE Transactions on Emerging Topics in Computing*. He serves/has served as General Co-Chair of 5GWN '17 and GameNets '16, Track Chair of ScalCom '15, Symposium Chair of GameNets '14, TPC Co-Chair of the INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks, General Co-Chair of the ICC 2018 (ICC 2017, GLOBECOM 2017) Workshop on 5G Ultra Dense Networks, and General Co-Chair of the GLOBECOM 2017 Workshop on LTE-U.He received the IEEE ComSoc Young Author Best Paper Award in 2017.

Yu Qiu (eeqiuyu@gmail.com) received her B.S. degree in electronic information engineering from Beijing University of Chemical Technology, China, in 2015. She is currently pursuing an M.S. degree at the Laboratory of Wireless Communications and Networks from the College of Information Science and Technology, Beijing University of Chemical Technology. Her research interests include 5G networks, fog RAN, mobility management, and resource allocation in wireless communications.

Xiaoli Chu [S'03, M'06, SM'15] (x.chu@sheffield.ac.uk) is a senior lecturer in the Department of Electronic and Electrical Engineering at the University of Sheffield, United Kingdom. She received her Ph.D. degree from Hong Kong University of Science and Technology in 2005. She has co-authored more than 100 peer-reviewed journal and conference papers. She is the lead editor/author of the book *Heterogeneous Cellular Networks — Theory, Simulation and Deployment* (Cambridge University Press, 2013). She is an Editor for *IEEE Wireless Communications Letters* and *IEEE Communications Letters*. She was Co-Chair of the Wireless Communications Symposium for ICC '15.

Keping Long [SM'06] (longkeping@ustb.edu.cn) received his M.S. and Ph.D. degrees at UESTC in 1995 and 1998, respectively. He worked as an associate professor at Beijing University of Posts and Telecommunications. From July 2001 to November 2002, he was a research fellow in the ARC Special Research Centre for Ultra Broadband Information Networks (CUBIN) at the University of Melbourne, Australia. He is now a professor and dean at the School of Computer and Communication Engineering, USTB. He is a member of the Editorial Committee of *Sciences in China Series F* and *China Communications*. He has also been a TPC and ISC member for COIN, IEEE IWCN, ICON, and APOC, and Organizing Co-Chair of of IWCMC '06, TPC Chair of COIN '05/'08, and TPC Co-Chair of COIN '08/'10, He was awarded the National Science Fund Award for Distinguished Young Scholars of China in 2007 and selected as the Chang Jiang Scholars Program Professor of China in 2008. His research interests are optical Internet technology, new generation network technology, wireless information networks, value-added service, and secure network technology. He has published over 200 papers, 20 keynotes, and invited talks.

Victor C. M. Leung [S'75, M'89, SM'97, F'03] (vleung@ece.ubc.ca) is a professor of electrical and computer engineering and holder of the TELUS Mobility Research Chair at UBC. He has co-authored more than 1000 technical papers in the area of wireless networks and mobile systems. He is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. He is a winner of the 2017 Canadian Award for Telecommunications Research and the 2017 IEEE ComSoc Fred W. Ellersick Prize.