REDESIGNING FRONTHAUL FOR NEXT-GENERATION NETWORKS: BEYOND BASEBAND SAMPLES AND POINT-TO-POINT LINKS

JINGCHU LIU, SHUGONG XU, SHENG ZHOU, AND ZHISHENG NIU

ABSTRACT

The fronthaul is an indispensable enabler for 5G networks. However, the classical fronthauling method demands large bandwidth, low latency, and tight synchronization from the transport network, and only allows for point-to-point logical topology. This greatly limits the usage of fronthaul in many 5G scenarios. In this article, we introduce a new perspective to understand and design fronthaul for next-generation wireless access. We allow the renovated fronthaul to transport information other than time-domain I/Q samples and to support logical topologies beyond point-to-point links. In this way, different function splitting schemes can be incorporated into the radio access network to satisfy the bandwidth and latency requirements of ultra-dense networks, control/data decoupling architectures, and delay-sensitive communications. At the same time, massive cooperation and device-centric networking could be effectively enabled with point-to-multipoint fronthaul transportation. We analyze three unique design requirements for the renovated fronthaul, including the ability to handle various payload traffic, support different logical topology, and provide differentiated latency guarantee. Following this analysis, we propose a reference architecture for designing the renovated fronthaul. The required functionalities are categorized into four logical layers and realized using novel technologies such as decoupled synchronization layer, packet switching, and session-based control. We also discuss some important future research issues.

INTRODUCTION

In recent years, cellular networks have witnessed a tremendous surge in data traffic, which is largely driven by the widespread adoption of smart devices such as smartphones and tablets [1]. While this trend will most likely continue in the foreseeable future, new challenges are also emerging with the proliferation of machine-type communications and real-time cloud services. In response to these predicted challenges, next-generation networks are envisioned to provide 1000× capacity, 100× data-rate, and 1 ms latency compared to fourth generation (4G) Long Term Evolution (LTE) systems [2]. The enabling technologies for such supreme performance are expected to include massive multiple-input multiple-output (MIMO), ultra-dense networking (UDN), as well as high-frequency spectrum.

Fronthaul (FH) is an important enabler for the deployment of these technologies in 5G networks. The term FH[1] has its root in the distributed base station (BS) architecture, in which the processing functions of a BS are split into two entities: the remote radio unit (RRU), which takes charge of radio processing and digital-analog conversion near the antennas, and the baseband unit (BBU), which handles digital baseband processing at another location. The classical form of FH refers to the point-to-point (P2P) link that transports time-domain complex baseband radio (a.k.a I/Q) samples between the corresponding RRU and BBU. Recently, FH also finds use in the novel centralized radio access network (C-RAN) architecture [3]. In C-RAN, time-domain I/Q samples are aggregated from scattered antenna sites to a central office for uplink (UL) processing or sent out in the opposite direction after downlink (DL) processing.

The dominant physical transmission technology for classical FH is digital radio over fiber (D-RoF). Although there are also several competing technologies such as analog RoF, the longer transportation range in C-RAN gives D-RoF great advantages due to its low signal deterioration. Various specifications have been formed to support the interoperability between FH products from different manufacturers. For example, the common public radio interface (CPRI) specification [4] covers layers 1 and 2 of FH. Its scope includes the physical topology, line data rates, framing format, and so on.

Although classical FH has been widely adopted in distributed BS and C-RAN, it will nevertheless face serious challenges in the face of 5G networks.

Jingchu Liu is with Tsinghua University and Intel Labs.

Shugong Xu is with Intel Collaborative Research Institute for Mobile Networking and Computing (ICRI-MNC).

Sheng Zhou, Zhisheng Niu are with Tsinghua University.

[1] This type of link got "front" in its name since it is closer to the network edge compared to backhaul (BH), which connects different BSs as well as BSs and core-network elements.

1536-1284/15/$25.00 © 2015 IEEE

1) Massive FH bandwidth requirements: The bandwidth requirements of a classical FH link are proportional to the product of radio bandwidth, number of antennas, and quantization resolution. To put this into perspective, a typical 20 MHz 4G LTE eNodeB with 8 antennas requires around 10 Gb/s FH bandwidth on UL or DL. Since the area density of antennas and the communications bandwidth are both expected to be enormous in 5G networks, the demand for FH bandwidth will become even more significant. Nevertheless, state-of-the-art compression techniques can only achieve at most 3× compression rate [5]. The most straightforward option for physical transport is dark fiber, in which one fiber core can only carry one FH link. However, this will result in enormous fiber resource consumption, making this scheme realistic only to operators with abundant fiber resources or in scenarios where fiber deployment is cheap. Another option is to multiplex FH links into a single fiber core using wavelength-division multiplexing (WDM), but WDM modules are still much more expensive than black and white modules.

2) Stringent latency constraints: Wireless signal processing often has stringent latency constraints. For example, the LTE hybrid automatic repeat request (HARQ) process leaves a latency budget of 3 ms for the decoding of each radio subframe. Because subframes need to first be transported from remote radio heads (RRHs) to BBUs before being processed, the transportation latency should also be counted into this latency budget. Excluding the portion necessary for signal processing (about 2.5 ms), there are only 300 to 500 μs left for FH transportation, ruling out any switching-based technologies that will incur excessive latency. Moreover, some 5G communication scenarios demand even stricter latency constraints: for sub-millisecond wireless access, the latency budget for FH transportation will be so small that any long-range transportation may be prohibited.

3) Tight synchronization requirements: Synchronization is essential for radio communication systems. But many RRHs cannot generate accurate clock by themselves, either because GPS receivers are too expensive to be integrated into RRH or due to satellite signal blockage in indoor environments. For this reason, FH must deliver synchronization information from BBUs to RRHs. In CPRI links, clock information is carried in the waveform (pulse edges) of the transported signal. But the underlying network infrastructure may introduce jitter to the waveform, leaving degraded communication performance. The importance of tight synchronization will be even greater in 5G networks because of the massive cooperation between access nodes. If cooperating access nodes have frequency offset, their transmitted signals will overlap at user equipment (UE), and will not be able to be separated and individually compensated, causing distorted beamforming patterns and degraded performance.

In this article, we propose a renovation of classical FH to address the above challenges. The renovated FH can transport intermediate processing information beyond time-domain I/Q samples. Also, it supports logical topologies beyond P2P links. The renovated FH can seamlessly incorporate different function splitting schemes and logical topologies, and can enable a number of key 5G concepts. We provide detailed analysis on three unique design requirements for the renovated FH:

- Handling various payload traffic
- Supporting flexible logical topology
- Providing differentiated latency guarantees

To facilitate efficient realization of the renovated FH network, we introduce a layered reference architecture and discuss how the layers may be realized using technologies such as decoupled synchronization, packet switching, and session-based management. Promising future research issues are also listed.

The rest of the article is organized as follows. We propose our renovation for classical FH and discuss how the renovated FH can enable some key 5G concepts. We analyze the three fundamental design requirements for the renovated FH. We then propose a layered reference architecture for realizing the renovated FH and introduce the enabling technologies. After that, we discuss some important future research issues. The article is then concluded.

Renovating FH

The classical understanding of FH is a P2P link between an RRH and BBU for transporting time-domain I/Q samples. In this section, we renovate this concept in two ways in order to address the challenges in 5G networks:

- FH should transport intermediate processing information other than time-domain I/Q samples.
- FH shall support not only P2P transportation but also point-to-multipoint networking.

We also illustrate how these new features could enable some promising 5G concepts.

Beyond Time-Domain I/Q Fronthauling

Recent research on BS function splitting revealed that allowing the transportation of intermediate processing information other than time-domain I/Q samples will help eliminate the bandwidth and latency bottlenecks of classical FH. The research on BS function splitting concerns the split of signal processing functions among different entities [6]. The classical RRH-BBU split is an extreme case, with minimal processing at RRHs. In contrast, alternative function splitting schemes place processing functions such as fast/inverse fast Fourier transform (FFT/iFFT), MIMO precoding/detection, modulation/demodulation, or even the whole physical layer (PHY) stack at RRHs. A trade-off between computational and networking resource can be achieved by using different function splitting schemes [7]. The optimal trade-off point should be decided considering the resources and constraints of specific scenarios. Analysis shows that the FH bandwidth requirements of alternative splitting schemes can be two orders of magnitude lower than the classical scheme, and the requirement on latency can also be relaxed [8].

The rationale behind these benefits can be explained as follows: signal processing functions in a sense append (extract) redundant informa-

The optimal trade-off point should be decided considering the resources and constraints of specific scenarios. Analysis shows that the FH bandwidth requirements of alternative splitting schemes can be two orders of magnitude lower than the classical scheme, and the requirement on latency can also be relaxed.

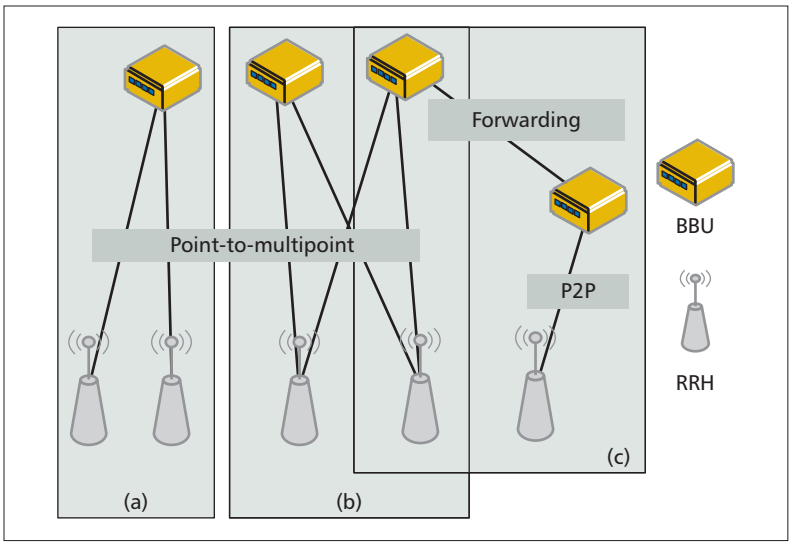

Figure 1. Possible logical topologies for FH networking.

tion to (from) the communications signals in order to combat channel deterioration. For example, modulation maps code-words (6 bits for 64-quadrature amplitude modulation, QAM) to complex constellation points (30 bits) so as to combat noises. Placing some of the processing functions at RRHs will reduce the amount of redundant information which need to be transported, and will in turn reduce the FH bandwidth requirements. As for latency, by moving latency-sensitive processing functions to remote sites, the fronthauling (FHing) latency will no longer be part of the total time budget, thus the latency requirement can be relaxed.

Since different function splitting schemes will result in different FHing payload, FH should be allowed to transport not only time-domain I/Q samples but also other intermediate processing information. Also, now that processing functions can be flexibly placed, the classical definitions of RRH and BBU should also be generalized: RRHs should be allowed to handle some extra processing functions, and BBUs need to be able to handle fewer processing functions.

From P2P FHing to FH Networking

Classical FH is in essence a logical P2P link, even though the underlying transport topology may be stars, rings, or chains. However, recent developments in cooperative communications put forward the need for point-to-multipoint logical topology. Massive cooperative is a main element in the technology evolution of RAN [9]. The information from (to) multiple RRHs is jointly processed to mitigate interference or increase cell throughputs. Typical cooperative processing schemes include joint transmission, joint reception, coordinated scheduling, and coordinated beamforming. As the next-generation networks become much denser, it is highly possible that cooperative processing will become prevalent in 5G networks.

Centralization is a possible way to implement cooperative processing. The processing information from multiple RRHs are aggregated through FH to one BBU, as in Fig 1a. In this way, the information of cooperating cells can be exchanged inside the BBU. When full centralization is too expensive, neighboring RRHs could form a cooperation cluster, and the processing information would only need to be aggregated within the cluster. In such a case, information must also be exchanged between clusters to facilitate communications on cluster edges. One way to exchange information between clusters is to establish FH links from one RRH to multiple BBU as in Fig. 1b. For this case, FFT/iFFT may need to be placed at RRHs so as the resource elements used for cooperative communication can be extracted and send to the cooperating cluster(s). Another way is to forward processing information from the BBU of one cluster to that of another cluster as in Fig. 1c.[2]

Enabling Key 5G Concepts

C/D Decoupling Architectures: Air-interface C/D decoupling architectures as proposed in the hyper-cellular network [10] and phantom cell [11] is a key concept for the management of dense small cells in 5G networks. In these architectures, the physical control channels are only transmitted by the control plane large cells, providing data plane small cells more chances to hibernate. In addition, the large control coverage also allows for better mobility management.

The renovated FH can support function splitting for more effective deployment of massive small data cells. The enormous FH bandwidth requirements of massive data cells can be satisfied by placing preprocessing functions at remote sites or utilizing a wireless FH link for the last hop; besides, the sparseness and burstiness of small cell traffic will be absorbed by statistical multiplexing, improving the utilization ratio of FH wireline resources. In both ways, the implementation flexibility of UDNs will be greatly improved in areas with limited fiber resource. Also, C/D decoupling inherently implies different FH payloads from control and data cells: control cells need to transport more control channel information, while data cells need to transport more data channel information. As illustrated in Fig. 2, this heterogeneous FHing demand can be satisfied more efficiently (compared to classical FH) by using FH links matched to the traffic patterns and delay requirements of control and data payloads, respectively.

Device-Centric Communications: Device-centric communications will be one disruptive technology direction for 5G networks [12]. In such a scenario, devices will simultaneously connect to multiple access nodes. This technology blurs the concept of a cell and breaks away from the previous cellular communication paradigm by which devices are connected to only one radio node at a time. To enable device-centric communications, the role of FH should also be changed from "transporting information for cells" to "networking information for devices." As illustrated in Fig. 3 (orange), the processing information of a user needs to be transported simultaneously from multiple RRHs to a BBU for joint processing. Note that different users may send their information to different BBUs. Moreover, the neighboring RRHs may change

[2] Forwarding information between BBUs on FH links may look similar to exchanging information between BSs through the X2 interface. However, because the main functionality of FH (i.e., networking intermediate processing information) is different from BH, the realization of their network interface shall be vastly different.

for users as they move around the network; thus, the set of serving FH links need to be adapted accordingly.

Latency-Sensitive Communications: Latency will be one of the key concerns in 5G networks. Although applications with normal delay requirements, such as common mobile Internet services, will still be needed, there will also be new applications like tactile Internet [13], which demands much shorter access latency (about 1 ms). The renovated FH can support these diverse latency requirements as in Fig. 3 (purple). For latency-sensitive applications, one option is to place the whole processing stack in order to avoid the transportation latency on FH; otherwise, low-latency FH links must be provisioned to ensure that the latency constraint is not violated. On the contrary, normal-latency applications may utilize FH links with looser latency guarantee. In more common scenarios, the latency requirements from different applications may vary significantly, and the renovated FH could provide a range of latency guarantee options for the resulting FH payloads.

Design Requirements

In this section, we analyze three design requirements that are unique to the renovated FH. These requirements include:

- Handling various payload traffic
- Supporting flexible networking topology
- Providing differentiated latency guarantee

Various Payload Traffic

A different function splitting scheme will introduce various bandwidth requirements for an FH network. As can be seen in Fig. 4, the baseline classical splitting scheme (time-domain I/Q FHing) consumes the most FH bandwidth. In comparison, the bandwidth requirements of alternative split schemes can be much lower than this baseline:

1. Low-pass-filtered I/Q FHing: Time-domain I/Q samples are low-pass-filtered before FH transportation. The blank guard band can be filtered out, roughly halving the FH bandwidth requirements.
2. Resource element extraction: All cell processing (e.g., FFT/IFFT) are placed at the RRH. Only the I/Q samples on active resource elements are extracted for FHing. In such a scheme, the instantaneous FHing rate is proportional to the actual radio resource usage, allowing the FH bandwidth requirements to further fall off in lightly loaded cells.
3. Modulation bits FHing: MIMO precoding/detection and modulation/demodulation functions are placed at the RRH; thus, the intermediate processing information to be transported is actually modulation information bits. Since modulation bits are much more compact representation than I/Q samples, the FH bandwidth is often at least one order of magnitude lower than baseline.

Note that these examples just illustrate the basic concept and are far from exhaustive. Other system configuration and function splitting schemes may result in different results. But the diversity in bandwidth requirements will be similar.

Besides bandwidth requirements, the FH traffic patterns of different splitting schemes will also manifest various randomness and periodicity (Fig. 4). The randomness is a consequence of cell load variations on one hand: the amount of radio resources under use varies as users come and leave, resulting in time-varying bandwidth requirements. On the other hand, it also results from the time-varying usage of modulation and coding schemes (MCSs) in response to channel fluctuations. Meanwhile, periodicity will also arise due to the transportation of periodic control signals like physical downlink control channel (PDCCH),

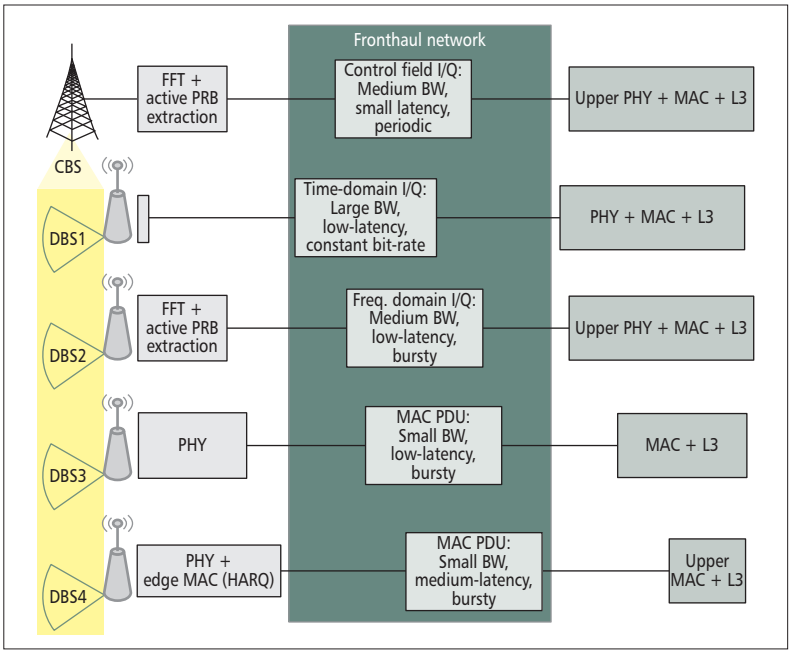

Figure 2. FH networking for C/D decoupled architecture and massive small cells. CBS and DBS can adopt different function splitting schemes (see function names in gray and green boxes) and will result in FH payload with different bandwidth, latency, and traffic pattern. (CBS: control base station, DBS: data base station, PRB: physical resource block, PHY: physical layer, MAC: medium access control layer, L3: layer 3 or network layer, PDU: packet data unit, BW: bandwidth).

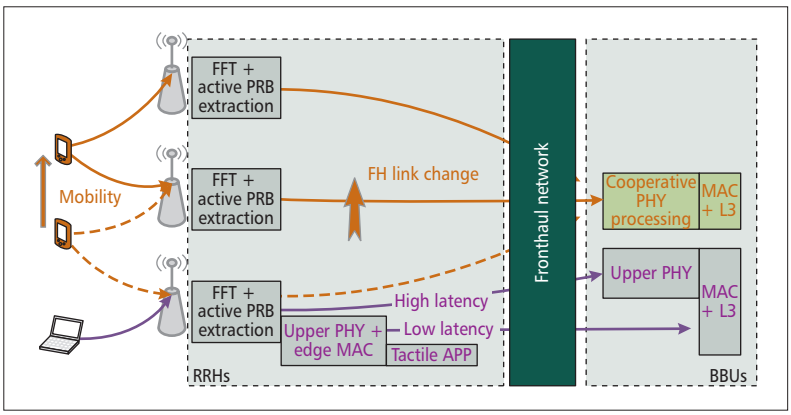

Figure 3. FH networking for device-centric (orange) and latency-sensitive communications (purple) (PRB: physical resource block, PHY: physical layer, MAC: medium access control layer, L3: layer 3 or network layer).

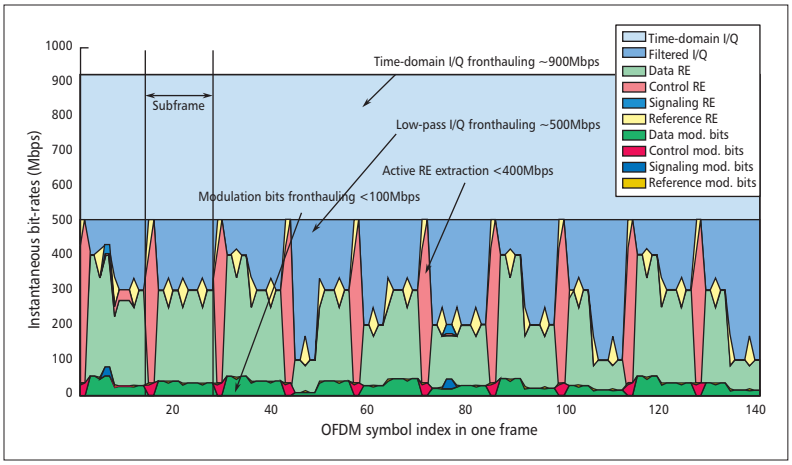

Figure 4. Instantaneous fronthauling bit-rates of different function splitting methods. Simulation parameters: 20MHz single cell and 10 UE, DL SISO transmission with link adaptation, sub-frame-level channel and user traffic variation.

physical random access channel (PRACH), as well as other overhead information.

These diverse bandwidth requirements and traffic patterns will render the traditional constant bit rate FHing technique, which only offers limited bandwidth and traffic pattern options, and is inefficient. For this reason, the renovated FH ought to provide sufficient flexibility to efficiently transport payloads with various bandwidths and traffic patterns.

Flexible Networking Topology

The logical topology of FH will be diversified in 5G networks. As mentioned previously, centralized cooperative processing requires an FH network to aggregate (distribute) information from (to) multiple RRHs to a BBU or transport information between BBUs. Besides, there may also be the need to transport information among RRHs when cooperative processing functions are placed at RRHs. Moreover, when the function splitting scheme is dynamically changed as in [7], the FH networking topology need to be adapted accordingly.

Aside from the support for different types of logical topologies, an optional design requirement is fine-grained topology management. There are two possible scenarios where the logical topology of FH links needs to be managed at sub-cell granularity. First, when different cells cooperate through FH links between RRHs, the information produced by control and signaling processing functions may still need to be transported to the BBU at the same time. In this case, the RRH-RRH and RRH-BBU links need to be managed separately. Second, the management granularity may need to be down to device-level granularity in device-centric communications: cooperative processing information for cell-edge users can be transported to the BBU, while the processing of other non-cooperative users can be placed locally at RRHs.

The renovated FH should have the ability to provide a very diverse collection of logical topologies, and provide fined-grained routing granularity for its payloads. Also, since the routing strategies will be tightly coupled with the selection of function splitting schemes, FH should provide the interface to enable joint design of function splitting and FH transportation.

Differentiated Latency Guarantee

Differentiated latency guarantee is needed for both the evolution of current networks and the development of 5G networks. In current networks, different function splitting schemes may result in different latency requirements. Take LTE, for example: if all processing functions that are below HARQ are conducted at RRHs, the latency requirement for FH can be relaxed to the 10 ms level (due to a PRACH latency constraint) [14]. For 5G networks, the demand for differentiated latency guarantee also comes from new applications. At one extreme, sub-millisecond wireless access is needed by future real-time applications such as tactile Internet. In such as case, either the transport latency on FH should be further reduced (e.g., to below 100 μs), or the whole processing stack should be placed at the RRH to avoid FH transportation. At the other extreme, there are also applications for which latency is not too much of a concern. A typical example is delay-tolerant machine-type communications. In such a case, the latency requirement on FH can be correspondingly relaxed. The relaxed latency requirements make payload scheduling and switching possible. Also, a large-delay transport network can be used in scenarios where no low-latency infrastructure is available. Between these two extreme cases, applications with intermediate latency requirements will also be possible. FH should have the ability to provision FH links with differentiated latency guarantee for different applications.

Realization Aspects

In this section, we discuss the realization aspects of the renovated FH. To facilitate discussion, we first propose a reference architecture for the renovated FH. As illustrated in Fig. 5, the physical layout of the proposed architecture is constructed of RRHs, BBUs, and FH switches. They are interconnected using physical links and form certain network topology, such as rings or chains.

The functionality of these physical network elements can be categorized into four logical layers. The foundation is a synchronization layer for distributing timing and clock references; in the middle are the payload layer for payload forwarding, and the control layer, which oversees payload forwarding and timing distribution; on the top sits the session layer, which presents logical FH links for applications. Next, we discuss some enabling technologies for these layers.

Decoupled Synchronization

As mentioned earlier, synchronization is essential for radio communication systems. In classical FH, synchronization signals are transported on FH links together with I/Q payload. This coupled mode works well on P2P links, but will create many ambiguities as the network topology becomes more complex. To address this difficulty, the synchronization functionality in the pro-

posed architecture is decoupled. The logical topology for timing and clock distribution is decided separately with payload transportation. Specifically:

- RRHs are slave nodes and synchronized to the signals from upstream FH nodes.
- BBUs receive synchronization signals from either other nodes or external timing sources.
- FH switches handle the selection, combination, and regeneration of timing and clock signals.

External timing sources can also be attached to FH switches as inputs. These processes are handled by dedicated circuitry like phase-locked loop (PLL) and timing processors, as well as specially designed message-based protocols and algorithms. A separate synchronization layer like this can also be found in other designs such as Synchronous Ethernet and IEEE 1588 [15].

Packet Switching

FH payload is transported in the FH network in the form of packets. As illustrated in the payload layer of Fig. 5, the data streams generated at RHHs or BBUs are first traffic-shaped by the regulator through buffering and framing, forming the transportation payloads. Each frame is then attached with a packet header, which indicates its FH link number, link counter number, and so on. The resulting packet is then handed to lower layer protocols for physical transportation. On the receiver side of the physical link, the FH switch first buffers the packet in the input buffer and extracts its header information. The local controller then processes the header and decides the scheduling and forwarding policy for this packet. The packet is then moved to the output buffers along with payloads from other FH links, and transported to the FH switch in the next hop. Once the packet has arrived at the destination end equipment, it is again buffered at the regulator, stripped of its header, and recovered into the payload data stream for further signal processing.

Packet-switched FH transportation has several advantages. First, a variable length packet can effectively handle the various payloads resulting from different function splitting schemes. Second, packetized payloads can be separately scheduled to increase the bandwidth utilization of a physical link under diverse payload bandwidth and traffic patterns, and to provide differentiated latency guarantee. Third, packets can be individually forwarded, forming the desired logical FH networking topology. Lastly, the FH payload can be dynamically switched to redundant physical paths under link or node failure, increasing the overall resilience of the network. Note that packet-based transportation of time-domain I/Q samples is being investigated by the IEEE 1904 Work Group under the name radio-(CPRI)-over-Ethernet. In contrast, our proposal also supports the transportation of information beyond I/Q samples and logical topologies other than P2P links.

Session-Based Control

The capacity of the FH network is presented in the form of FH sessions. An FH session stands for one stream of FH payloads between a pair of end equipment and having certain bandwidth and latency guarantee. The payloads in the same session share the same physical path in the network. Each session is set up and torn down separately in the network through the control layer. The control layer calculates the best transportation path and configures the local control entities along the path.

This session-based control is realized using virtual circuit switching, which establishes virtual connections in a packet-switched network. Connect-oriented switching has a number of advantages over its connectionless counterparts, such as better bandwidth and latency guarantee, lower switching overhead, and higher switching speed. This kind of technology has been widely used in major protocols like multiprotocol label switching (MPLS). These features are all favorable to FH transportation, since FH payloads

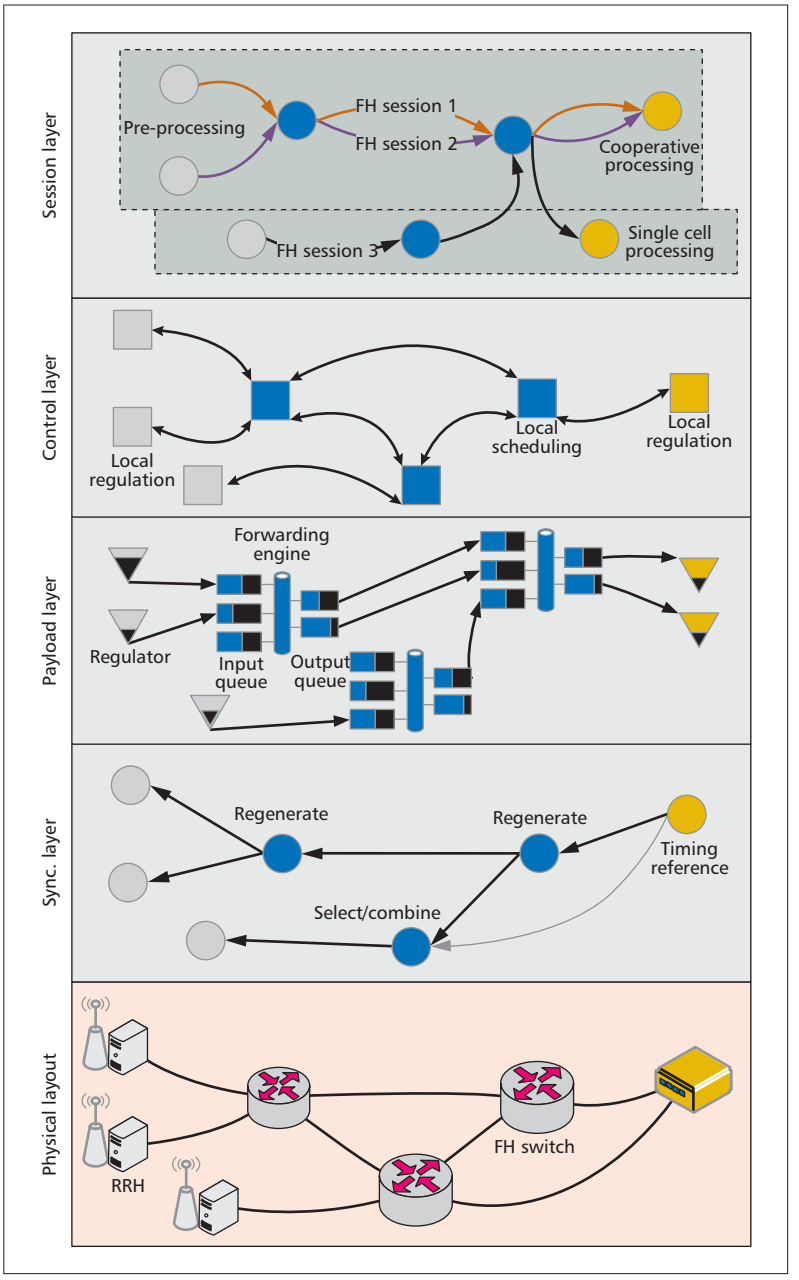

Figure 5. Reference architecture.

To provide uniform QoS guarantees along the FH link, performance monitoring, reporting, and negotiation capabilities must be designed at segment gateways, and the QoS differences of different segments should be considered by the FH when scheduling and forwarding payloads.

may have different bandwidth and latency requirements, but also demand high switching capacity and low switching delay.

Future Research Issues

In this section, we discuss some important research issues for the renovated FH.

Overhead Analysis and Reduction

The renovated FH network provides great flexibility over payload traffic, topology, and latency guarantee. However, along with the great flexibility comes increased overhead. First, the packet headers do not carry any payload information and will thus reduce the overall bandwidth efficiency. Second, header processing and payload scheduling will introduce additional latency to FH transportation and require more computational resources. Moreover, fine-grained FH session management will bring more negotiation and signaling overhead.

These overheads should be thoroughly quantified and analyzed in the future in order to identify fundamental design trade-offs, which can be used for overhead reduction. For example, a shorter header may reduce the control information carried in each packet and hurt flexibility, but it will increase bandwidth efficiency and speed up the control processing pipeline. Likewise, restrictions on available payload length options may reduce the bandwidth efficiency, but will provide better latency guarantee. Also, managing FH sessions in a group will reduce flexibility, but at the same time the overall signaling overhead. Efficient implementation will have to search for an optimal trade-off point.

Software-Defined Networking

Software-defined networking (SDN) is a novel paradigm to enhance the efficiency, flexibility, and management of networks. This concept can also be introduced into the renovated FH. Note that in the reference architecture, the control layer does not mandate a distributed implementation. A centralized controller could be introduced into this layer to coordinate payload scheduling and forwarding as well as timing distribution based on global network information. However, in order to introduce SDN into FH networks, the common issues with SDN must be addressed. For example, scalability is a major concern in any architectures with centralized control. The central controller may be overloaded by massive control signaling demands and complex control algorithms, resulting in large latency or packet drop. This issue calls for novel control plane designs such as hierarchical control design. Also, centralized control facilitates SDN operation. Applications can access the capabilities of the proposed architecture by calling application programming interfaces (APIs) on the central controller. However, the APIs must be designed according to the special needs of FH networking.

Heterogeneous FH Links

The available link technologies for FH will continue to evolve and blossom in the future. Optical modules will have higher rate and lower price thanks to the development of technologies like silicon photonics. Wireless link technologies such as mmWave and free space optics (FSO) are also likely to become mature enough to transport FH payloads as well as control and synchronization signals. Both types of link technologies have their own pros and cons. Therefore, it is important for the renovated FH to utilize heterogeneous FH links in a synergistic manner. For example, wireless links could provide last-hop FH access, while optical transport can be used to aggregate last-hop access links.

These different network segments may employ different protocols, so segment gateways need to be designed across which to transparently deliver payloads. Another issue is the non-uniform link quality of service (QoS) among segments. For example, public networks often have less bandwidth and larger latency than private networks. To provide uniform QoS guarantees along the FH link, performance monitoring, reporting, and negotiation capabilities must be designed at segment gateways, and the QoS differences of different segments should be considered by the FH when scheduling and forwarding payloads. The feasibility and best practice of synergetic operations of heterogeneous FH links require further investigation.

Networking-Processing Co-Design

In traditional RAN, the performance of wireless communication is decoupled with computational and wireline resources. Each BS is equipped with enough computational resources to handle signal processing on its own, even in peak hours. And the information exchange demand between BSs, and between BSs and the core network elements is minimized since all signal redundancy is removed through local processing.

However, the renovated FH allows for the co-design of processing and networking. This is because the processing information can now be transported to arbitrary places for processing, allowing for trade-off between computational, wireline, and radio resources. Possible trade-offs include processing consolidation (wireline–, computation+),[3] FH compression and function splitting (computation–, wireline+), centralized cooperative processing (computation–, wireline–, radio+), wireless FH (radio–, computation–, wireline+). Preliminary examples of such trade-offs have emerged in the form of fully centralized processing in C-RAN. However, more alternative forms of trade-off should be investigated to address the practical challenges of cloud-based radio access, such as high FH bandwidth requirements and energy consumption.

Conclusion

In this article, we renovate the classical FH to address the challenges in 5G networks. The renovated FH can transport intermediate processing information beyond time-domain I/Q samples, and allows for logical topologies other than P2P links. In this way, different function splitting schemes and logical topologies could be employed to enable key 5G concepts without violating the bandwidth and latency constraints. With respect to the renovated FH, we highlight

[3] + means the trade-off is in favor of this kind of resource, while – means the trade-off acts against this kind of resource. If not mentioned, this kind of resource is not affected by the trade-off.

the three unique design requirements, which are handling various payload traffic, supporting flexible logical topology, and providing differentiated latency guarantees. We also provide a layered reference architecture for realizing the renovated FH, and discuss key enabling technologies and important future research issues.

Acknowledgment

This work is sponsored in part by the National Basic Research Program of China (No. 2012CB316001), and the National Natural Science Foundation of China (NSFC) under grant No. 61201191 and 61401250, the Creative Research Groups of NSFC (No. 61321061), the Sino-Finnish Joint Research Program of NSFC (No. 61461136004), and the Intel Collaborative Research Institute for Mobile Networking and Computing.

References

[1] Cisco, "Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018," 2013; http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html.
[2] J. Andrews *et al.*, "What Will 5G Be?" *IEEE JSAC*, vol. 32, no. 6, June 2014, pp. 1065–82.
[3] China Mobile Research Inst., "C-RAN: The Road towards Green RAN," v. 3.0, June 2014; http://labs.chinamobile.com/cran/wp-content/uploads/2014/06/20140613-C-RAN-WP-3.0.pdf
[4] CPRI Cooperation CPRI Dpecification v6.0: Interface specification, Aug. 2013.
[5] D. Samardzija *et al.*, "Compressed Transport of Baseband Signals in Radio Access Networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, Sept. 2012, pp. 3216–25.
[6] NGMN Alliance, "Suggestions on Potential Solutions to C-RAN," v. 4.0, Jan. 2013; http://ngmn.org/uploads/media/NGMN_CRAN_Suggestions_on_Potential_Solutions_to_CRAN.pdf.
[7] J. Liu *et al.*, "Graph-Based Framework for Flexible Baseband Function Splitting and Placement in C-RAN," *IEEE ICC '15*, June 2015.
[8]. U. Dotsch *et al.*, "Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE," *Bell Labs Tech. J.,* vol. 18, no. 1, 2013, pp. 105–28.
[9]. R. Irmer *et al.*, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Commun. Mag.*, vol. 49, no. 2, Feb. 2011, pp. 102–11.
[10]. Z. Niu *et al.*, "Energy Efficiency and Resource Optimized Hyper-Cellular Mobile Communication System Architecture and Its Technical Challenges," *Science China: Info. Science*, vol. 42, no. 10, 2012, (in Chinese), pp. 1191–1203.
[11]. H. Ishii, Y. Kishiyama, and H. Takahashi, "A Novel Architecture for LTE-B :C-Plane/U-Plane Split and Phantom Cell Concept," *IEEE GLOBECOM '12*, Dec. 2012.
[12]. F. Boccardi *et al.*, "Five Disruptive Technology Directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 74–80.
[13]. G. Fettweis and S. Alamouti, "5G: Personal Mobile Internet Beyond What Cellular Did to Telephony," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 140–45.
[14]. D. Wubben *et al.*, "Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible Centralization through Cloud-RAN," *IEEE Signal Processing Mag.*, vol. 31, no. 6, Nov. 2014, pp. 35–44.
[15]. S. Rodrigues, "IEEE-1588 and Synchronous Ethernet in telecom," *IEEE Int'l. Symp. Precision Clock Synchronization for Measurement, Control and Communication 2007*, Oct. 2007, pp. 138–42.

Biographies

JINGCHU LIU [S] (liu-jc12@mails.tsinghua.edu.cn) received his B.S. degree in electronic engineering from Tsinghua University, China, in 2012. He is currently a Ph.D. student at the Department of Electronic Engineering, Tsinghua University. His research interests include cloud-based wireless networks, network data analytics, and green wireless communications.

SHUGONG XU [SM] (shugong.xu@intel.com) graduated from Wuhan University, China, in 1990, and received his Master and Ph.D. degrees from Huazhong University of Science and Technology, China, in 1993 and 1996, respectively. He is currently the center Director and Intel Principal Investigator of the Intel Collaborative Research Institute for Mobile Networking and Computing (ICRI-MNC). Before joining Intel in September 2013, he was a research director and principal scientist at the Communication Technologies Laboratory, Huawei Technologies. Among his responsibilities at Huawei, he founded and directed Huawei's green radio research program, Green Radio Excellence in Architecture and Technologies (GREAT). He was also the Chief Scientist and PI for the China National 863 project on End-to-End Energy Efficient Networks. Prior to joining Huawei in 2008, he was with Sharp Laboratories of America as a senior research scientist. Shugong has published more than 60 peer-reviewed research papers in top international conferences and journals. One of his most referenced papers has more than 1200 Google Scholar citations, in which the findings were among the major triggers for the research and standardization of IEEE 802.11S. He has more than 20 U.S. patents granted. His recent research interests include mobile networking and computing, next generation wireless communication platforms, network intelligence and SDN/NFV, green communication, etc.

SHENG ZHOU [M] (sheng.zhou@tsinghua.edu.cn) received his B.S. and Ph.D. degrees in electronic engineering from Tsinghua University in 2005 and 2011, respectively. He is currently an assistant professor with the Electronic Engineering Department, Tsinghua University. From January to June 2010, he was a visiting student at the Wireless System Lab, Electrical Engineering Department, Stanford University, California. From November 2014 to January 2015, he was a visiting researcher in the Central Research Lab of Hitachi Ltd., Japan. His research interests include cross-layer design for multiple antenna systems, cooperative transmission in cellular systems, and green wireless communications.

ZHISHENG NIU [F] (niuzhs@tsinghua.edu.cn) graduated from Beijing Jiaotong University, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. During 1992–1994, he worked for Fujitsu Laboratories Ltd., Japan, and in 1994 he joined Tsinghua University, where he is now a professor with the Department of Electronic Engineering and deputy dean of the School of Information Science and Technology. He is also a guest chair professor of Shandong University, China. His major research interests include queueing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communication and networks. He is a Fellow of IEICE, a Distinguished Lecturer (2012–2015) and Chair of the Emerging Technology Committee (2014–2015) of IEEE Communication Society, and a Distinguished Lecturer (2014–2016) of the IEEE Vehicular Technologies Society.

The renovated FH can transport intermediate processing information beyond time-domain I/Q samples, and allows for logical topologies other than P2P links. In this way, different function splitting schemes and logical topologies could be employed to enable key 5G concepts without violating the bandwidth and latency constraints.