

Virtualized Cloud Radio Access Network for 5G Transport

Xinbo Wang, University of California, Davis

Cicek Cavdar, KTH Royal Institute of Technology, Sweden

Lin Wang, University of California, Davis

Massimo Tornatore, Politecnico di Milano, Italy

Hwan-Seok Chung, Han-Hyub Lee, Soo-Myung Park, ETRI Korea

Biswanath Mukherjee, University of California, Davis

Abstract

Current Radio Access Networks (RANs) need to evolve to handle diverse service requirements coming from the growing number of connected devices and increasing data rates for the upcoming 5G era. Incremental improvements on traditional distributed RANs cannot satisfy these requirements, so the novel and disruptive concept of cloud-RAN (CRAN) has been proposed to decouple digital units (DUs) and radio units (RUs) of base stations (BSs), and centralize DUs into a central office, where virtualization and cloud computing technologies are leveraged to move DUs in the “cloud”. However, separating RUs and DUs requires low-latency and high-bandwidth connectivity links, called “fronthaul”, as opposed to traditional backhaul links. Hence, design of the 5G transport network, i.e., the part of the network that carries mobile data traffic between BSs and the core network and data centers, is key to meet the new 5G mobile service requirements and effectively transport the fronthaul traffic. Today, consensus has not yet been achieved on how the fronthaul traffic will be transported between RUs and DUs, and how virtualization of network resources will occur from radio network segment to the centralized baseband processing units. In this article, we present a new 5G architecture called virtualized-CRAN (V-CRAN) moving towards a cell-less 5G network architecture. We leverage the concept of a “virtualized-BS” (V-BS) that can be optimally formed by exploiting several enabling technologies such as software-defined radio (SDR) and Coordinated Multi-Point (CoMP) Transmission/Reception. V-BS can be formed on a per-cell basis or per-user basis by allocating virtualized resources on demand. For the fronthaul solution, our approach exploits the passive optical network (PON), where a wavelength can be dynamically assigned and shared to form “virtualized” PON (VPON). Several use cases of V-CRAN are presented to show how network architecture evolution can enhance system throughput, energy efficiency, and mobility management.

1. Introduction

5G networks are envisioned to support 1000x more traffic than today; however, the cost and energy consumption should support affordable services and sustainable growth [1]. 5G RAN design is not only about traffic increase, but also about supporting a large variety of services and devices with unprecedented quality of experience (QoE) [1]. Considerable growth of traffic demand and service types pose serious issues of scalability and management for network operators. Even though mobile network capacity can be enhanced by (1) densification, i.e., by deploying more radio units and adding another capacity layer with small cells and (2) by usage of larger spectral resources, traditional RANs, based on a distributed architecture, fall short in satisfying these requirements. This calls for a new centralized RAN architecture where processing resources are shared to minimize cost and enable advanced coordination techniques between base stations (BSs). Distributed and centralized RANs and the driver for this evolution are explained below.

1.1. Distributed RAN (DRAN)

In DRAN, each BS has two components, co-located in the same cell site: (1) Digital Unit (DU) or Baseband Unit, and (2) Radio Unit (RU) or Remote Radio Head. DU is responsible for baseband processing, while RU is responsible for transmitting/receiving and digitizing radio signals. The communication channel between each BS and the core network is referred as backhaul. In DRAN, one way to increase capacity is to densify the network, but cost also increases with additional BSs deployed, as a DU dedicatedly

serves its associated RU, and processing resources of a DU cannot be shared among other RUs. Another way to increase capacity is to use advanced transmission technologies such as Coordinated Multi-Point (CoMP) Transmission/Reception, which is a suite of radio transmission techniques to reduce interference and increase system throughput. However, this **applies** stringent delay constraints for control and signaling to guarantee on-time coordination between BSs [2]. But the processing resources of distributed BSs in DRAN are not designed for CoMP, and signaling exchanges undergo long delay (4-15 ms) over backhaul links connecting BSs to the core network [2]. Therefore, DRAN does not represent a future-proof and scalable solution for 5G.

1.2. Cloud-RAN (CRAN)

Recently, CRAN architecture has been proposed [3], which moves the DU from distributed BSs to a centralized location, serving a large group of RUs. Centralization can be done by “DU hoteling”, where DUs are co-located but remain separate, and each DU is designated to a specific RU; or by “DU pooling”, where DUs can be shared by different RUs and serve as a resource pool. “DU pooling” can then evolve to “DU-cloud” if pooled DUs are implemented on general-purpose servers that are flexibly configured [4].

DUs and RUs need to exchange signals and data, which requires a transport network with high bandwidth, short latency, good jitter performance, etc. This part of the network is called “fronthaul”. Optical channels, over individual fibers or wavelengths, are viable options. However, high-capacity optical channels need to be shared by different BSs adaptively. A framework for implementing time-wavelength division multiplexed (T-WDM) channels over a common fiber network is provided by the PON architecture, and it is known as TWDM-PON [5]. **CRAN over TWDM-PON was first proposed in [6], and [7] utilizes data-compression techniques to reduce fronthaul bandwidth for a TDM-PON-based-CRAN (both these works only focus on the fronthaul of CRAN).** With upcoming 5G technologies such as mm-wave and Massive MIMO, new trends are developing in terms of different DU and RU splits (supported by “midhaul” [8]) and new interfaces enabling packet-switching-based fronthaul [9].

Cloud RAN leverages recent trends of virtualizing network functions and providing them on demand, as enabled by technologies such as software-defined networking (SDN) and network function virtualization (NFV). Thus, to change the way mobile services are provided and fulfill the requirements of 5G infrastructure for “anything-as-a-service” [1], a joint framework is needed to enable full virtualization of resources in the 5G transport network starting from DU-cloud to fronthaul and even BSs, as discussed below.

1.3. Virtualized-CRAN

In this article, we present a new 5G network architecture called virtualized-CRAN (V-CRAN), and its building blocks and design principles. We introduce novel concepts such as (1) Virtualized PON (VPON) and (2) Virtualized BS (V-BS); and we show how virtualization and adaptive allocation of resources will take place from DU-cloud to RUs.

The first building block is VPON, a virtualized communication channel over a wavelength between many RUs and a DU as in an independent PON. TWDM-PON can provide many such VPONs. VPON can associate geographically-adjacent RUs with the same DU that can have global information about these RUs and coordinate them with specialized hardware/software for CoMP, and VPON can provide dedicated transport of layer-1 digitized signals for a DU, considering the latency budget of CoMP. So, V-CRAN can ease the implementation of CoMP through VPON formation (explained in Section 2).

A V-BS represents a combination of processing resources in DU-cloud, shared VPON in fronthaul, and a set of RUs (also radio resources of RU) in radio access layer. For V-BS, processing resources in a DU are virtualized as functional entities that can be migrated within the DU-cloud. We can form a V-BS for a RU even for a particular user on demand. When a user is mobile, CoMP provides seamless communication by re-forming dynamic clusters of RUs that can jointly transmit signal to the user. For resource allocation in V-CRAN, V-BS design principles need a cross-layer optimization framework, which assigns resources in an end-to-end manner, i.e., by allocating (1) DU processing resources, (2) fronthaul transmission resources, and (3) radio resources for each user. This requires that V-BS design jointly consider constraints in different segments of V-CRAN, e.g. interference avoidance in radio network, fronthaul capacity in fronthaul, and processing capacity in DU cloud.

The contributions of this paper are three-fold: 1) we describe our proposed vision for a V-CRAN architecture that virtualizes DU cloud, fronthaul, and radio sites in a holistic view; 2) we discuss how to use the concepts of VPON and VBS to jointly optimize allocation of resources in different segments of the V-CRAN; and 3) we quantitatively evaluate the benefits of V-CRAN in terms of throughput enhancement, energy saving, and handover reduction.

2. Virtualized Passive Optical Network (VPON)

2.1. Physical Architecture Overview

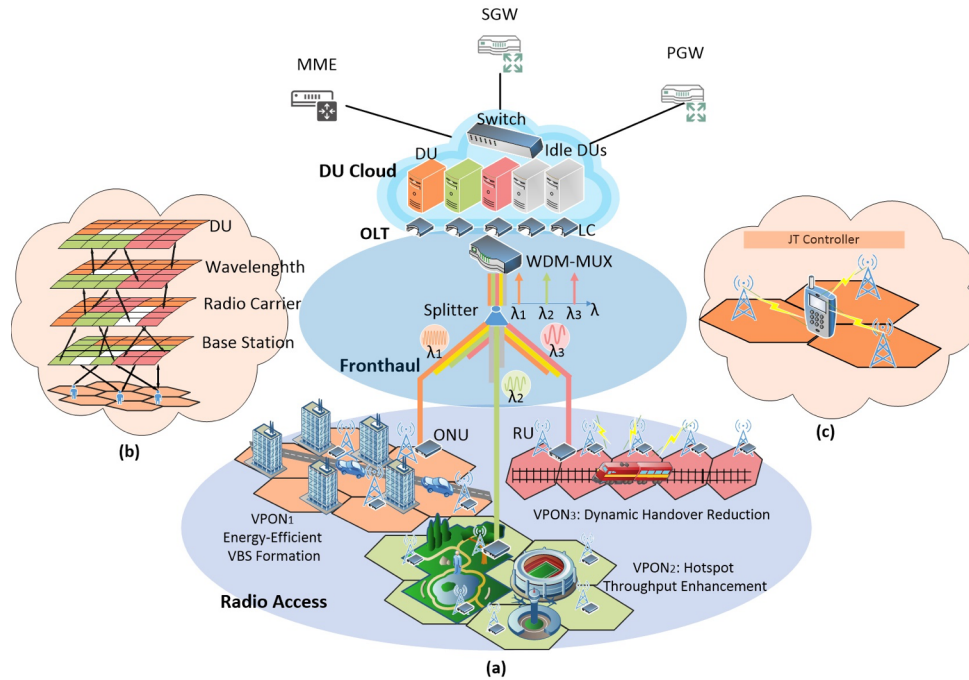


Figure 1. a) Virtualized-CRAN Architecture. b) Illustration of joint optimization of resources. c) Illustration of joint transmission.

*MME(mobility management entity), SGW (service gateway) and PGW (PDN gateway) are functions of mobile core networks. They can be deployed in mobile core or V-CRAN.

We illustrate the V-CRAN architecture in Fig. 1a. In DU-cloud, commercial servers can be customized to provide real-time baseband processing (and other layer2/layer3) functions; and such a server can play the role of a DU [8]. DUs are inter-connected by a high-speed Layer-2 switch, which exchanges signaling and data among DUs.

TWDM-PON is used in fronthaul because it can satisfy the stringent delay requirements of fronthaul segment, and provide abundant bandwidth, at low cost and energy consumption [5]. TWDM-PON consists of an optical line terminal (OLT) and many optical network units (ONUs). OLT is collocated with DU-cloud and provides each DU with an optical transceiver and a linecard (LC) that can deliver traffic over a wavelength and provide optical-electrical conversion. LCs are connected with a WDM multiplexer/de-multiplexer (MUX) that can separate traffic on a per-wavelength basis. The passive splitter located remotely from DU-cloud can branch the fiber to enlarge the coverage of a TWDM-PON. At the end of a fiber, there is an ONU, which is equipped with a reconfigurable transceiver to tune its transmission wavelength. An ONU is collocated with a RU as an agent of the fronthaul optical network. **Note that, with the densification of BSs and increasing traffic load, signaling (e.g., for handover) between the evolved packet core (EPC) and BSs in LTE becomes a challenging problem. V-CRAN can ease the pain by minimizing the number of handovers and reducing the signaling burden due to base-station coordination**

2.2. Wavelength and DU sharing

The grouped ONUs share the same wavelength and form a VPON. Each VPON can work independently, using, e.g., Ethernet or other TDM-based approaches. The control software in DU cloud can reconfigure the VPONs by instructing ONUs to retune their transceivers to the operating wavelength of the target VPON, which incurs extra delay due to the tuning time and signaling exchange. VPON formation is envisioned to be carried out infrequently (e.g., at time scale of minutes or hours) and hence not pose excessive burden to the network control system.

Traffic generated in a RU's coverage area (i.e., inside a "cell" or a sector in a cell) can be grouped into a VPON to share a wavelength's bandwidth. Traffic load in RAN is dynamic and has temporal pattern [10], e.g., alternating busy and idle periods, and spatial pattern, e.g., with shifted windows of busy hours for urban and residential areas. During busy hours of a day when cells are highly loaded, each cell can be assigned a dedicated VPON. During idle hours, it is desirable to group cells onto fewer VPONs, so that idle VPONs and their associated DUs, as well as optical components and interfaces, can be turned off (or put into sleep mode). Besides, DUs and VPON that were assigned to an area undergoing idle hours can be shared by areas currently experiencing busy hours. Thus, the network can be utilized more efficiently and operational costs are decreased. When traffic

load rises and falls in CRAN, VPONs can be reformed accordingly, e.g., as shown in Fig 1a. VPON₁ (orange color) can be formed to serve the urban area during idle hours.

Any load consolidation/balancing technique can serve this purpose. A layer-2 switch with enough ports can redirect each frame from any cell to any DU. But when the RAN enlarges and is densified with more cells, the complexity of hardware and control software will grow, and switch latency may negatively impact the performance of base-station coordination algorithms such as CoMP. If traffic load from a cell tends to be stable over sustained periods (e.g., minutes), load consolidation/balancing should be done on a per-cell basis in layer-1, instead of a per-frame basis in layer-2. V-CRAN performs this operation in optical domain on the fronthaul. This design can reduce the DU-cloud's complexity and shorten the latency, considering CoMP's rigid latency requirement (typically 1 ms or less) [2].

VPON formation can be formulated as an optimization problem (and solved by Integer Linear Program (ILP) [11]) to minimize usage of wavelengths and DUs. The solution provides (a) wavelength assignment of each ONU and (b) association of each RU with a DU using the following design principles:

- To associate a RU to a DU, the ONU equipped with the RU must be tuned on the same wavelength of the DU, which is called **wavelength-uniformity constraint**.
- Total traffic load from all RUs belonging to a VPON must not exceed the capacity of a wavelength or a DU.

2.3. Base-Station Coordination

VPON formation enables not only resource sharing, but also base-station coordination. RUs located in an area can be grouped into a VPON and controlled by the same DU. The whole radio access area can be partitioned into many service areas by formation of VPONs (see Fig. 1a). Within each service area, V-CRAN can implement CoMP to improve RAN coverage, bit rate, and throughput. Attractive CoMP techniques providing the highest gains are inter-cell interference cancellation (ICIC) and joint transmission (JT). ICIC reduces interference by allocating different physical resource blocks to users in neighboring cells to avoid overlapping of RBs while scheduling users (spectrum band of a RU is divided into continuous physical resource blocks (RB) with fixed size). JT is applied in contexts where multiple adjacent BSs cooperate to transmit and receive signals for a user over the identical RB, so that interference can be converted to useful information. ICIC needs global information about occupation status of RBs in different cells. JT requires not only global information, but also heavy processing resources and short transmission latency to achieve high performance. Also, data and signaling information for scheduling should be duplicated and delivered to all coordinating BSs before the scheduling decision is taken [2]. DRAN falls short of satisfying these requirements of JT as discussed in Section 1. Such delay can be shortened in V-CRAN for two reasons. First, TWDM-PON can provide multiple independent VPONs with fast transport and processing of data between RUs and DU. Second, as shown in Fig. 1c, each DU has global information of all RUs associated with it and is equipped with dedicated hardware/software as a JT controller to provide JT computation.

VPON can also be formed with the objective of maximizing base-station coordination by modifying the optimization problem stated in Section 2.2. By solving it, we can split the radio access area into multiple service areas, each served by a cluster of RUs and an associated VPON. A modified optimization problem can also minimize the usage of wavelengths and DUs. **Note that pooling more RUs in a VPON can achieve more coordination gains, by distributing functions to RUs or increasing the wavelength bandwidth, but gains tends to saturate due to other constraints, e.g., DU's baseband processing capability and distance between RU and the users.**

3. Virtual-Base Station (V-BS)

Depending on the objective, a V-BS formation can be classified as cell-centric or user-centric.

3.1. Cell-Centric V-BS Formation

We can allocate "just enough" virtualized resources for each cell to form a cell-centric V-BS, comprised of a shared VPON and virtualized functional entities of DU, as shown by red dotted lines in Fig. 2 to illustrate a V-BS for Cell₂. Two types of functional entities exist in a DU: cell processing (CP) and user processing (UP) [13]. A DU can set up a CP for each cell associated with it (see Fig. 2). A CP provides part of baseband processing, multiplexes/de-multiplexes user traffic from a cell, and processes cell-control messages. A cell can change its serving DU through *CP shifting*, which first retunes the wavelength of the ONU to change the serving VPON, then sets up a new CP in the destination DU, and finally reclaims the old one in the source DU. A UP provides remaining baseband processing and customized service for each user. A UP can be redirected from one DU to another by *UP*

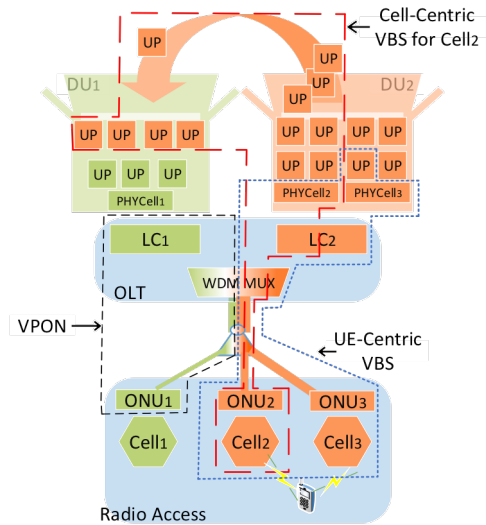


Figure 2. Illustration of VPON and V-BS.

redirection, in which an overloaded DU redirects some UPs to DUs with extra processing resources through active ports of the internal switch. For example, in Fig. 2, when DU₂ cannot accommodate all UPs of Cell₂, it can redirect some UPs to DU₁. So, V-BSs form a cell-less architecture, where there is no dedicated wavelength in fronthaul and functional entities in DU for a specific RU.

The cell-centric V-BS formation problem can be modeled by following the principles described for the optimization problem in Section 2.2 with the objective to minimize the usage of wavelengths, CPs and UPs in DU, and switch ports [11]. By solving this problem, besides the findings in Section 2.2, we can also determine (a) for each cell, which DU to select to assign a CP; (b) for each user, which DU to select to assign a UP; and c) number of active switch ports. So, we add the following design principles:

- For a cell, a CP must be set up in the DU that the RU of the cell is associated with.
- If a DU is heavily loaded, UPs can be redirected and set up in other DUs, but the source and destination DUs of UP redirection must occupy active ports of the internal switch.

3.2. User-Centric V-BS Formation

We can also form user-centric V-BS for each user, comprised of a group of RUs, a shared VPON, and functional entities in DU, as shown by blue dotted lines in Fig. 2. User-centric V-BS formation assisted by ICIC and JT can enhance the throughput and QoE for users. To implement JT, a set of RUs that can provide good signals for a user needs to be selected from neighboring cells into the “JT-set”. Although throughput of a user can be enhanced by selecting more RUs into the JT-set to form a “bigger” V-BS, it consumes more RBs and increase interference for other users, whose throughput might be jeopardized. ICIC can alleviate this problem by assigning different RBs to users in neighboring cells; but when traffic load is heavy, it is hard to assign RBs without conflict due to diminishing available RBs.

V-BS can be jointly formed with VPON with the objective of maximizing total throughput of all users in V-CRAN [12]. By solving the model using constraint programming [14], we obtain VPON formation and RU-DU association, and a) which RB is assigned to a user, b) which RU provides JT service for a user, and c) which RU sends interfering signal to a user. V-BS formation should follow all design principles described in Section 2, especially wavelength-uniformity constraint, which requires RUs in JT-set to be in the same VPON so that they can be controlled by the same JT controller in DU. There are additional design principles, as follows:

- All RUs in JT-set must transmit common data to a user over the identical RB, which is called **resource-block-continuity constraint**. This can significantly simplify the transceiver required by user.
- Inside the DU, CP of cells in the JT-set and UP of the user must be set up together so that JT data and control messages can be exchanged and processed fast.
- To avoid providing unfair JT services, JT-set of a user can only be selected from cells that the user can receive signals from, and must contain as least the host cell, which is the nearest RU with the user.
- For those RUs that can transmit signals to a user but not selected in its JT-set, they should avoid allocating the RB that the user is occupying to reduce interference.
- Number of RBs of a RU allocated to users must not exceed the capacity of RU’s bandwidth resources.

- A RB at a RU cannot be assigned to more than one user at a time.

3.3. Moving-User-Centric V-BS Formation

When a user is moving, V-BS can be dynamically formed surrounding the user, delivering data to the area at the user's arrival. Typically, when a user is going across the cell edge, signal strength from the serving cell diminishes. If the signal strength is lower than a threshold and there is another target cell that can provide a stronger signal, a handover is triggered for the user between serving cell and target cell [15]. When the user is moving within the service area served by a common VPON but traversing a cell edge, V-BS can be formed to provide a strong-enough signal for the user so that the handover will not be triggered.

So, we add the following design principles for the moving-user-centric V-BS formation problem (besides those described in Sections 2 and 3.2):

- V-BS cannot be formed for a mobile user if the wavelength-uniformity constraint is violated, when it goes across the boundary of two VPONs. A handover is needed to set up a new UP in destination DU, migrate information of the user, and reclaim the old UP in source DU, which we call "inter-VPON" handover.
- V-BS cannot be formed for a mobile user moving within a VPON if the resource-block-continuity constraint is violated, even though it is moving within a VPON. A handover is needed to assign a new RB for the user, which we call "intra-VPON" handover.

In V-CRAN, both types of handovers are allowed within DU-cloud through internal communication, which are complex signaling exchanges between BSs. Handover latency and failure rate can be reduced also.

4. Case Studies

Now, we present three use cases of V-CRAN, and demonstrate how V-CRAN can outperform the two reference architectures--DRAN and CRAN--with respect to throughput, energy efficiency, and mobility management, and where the enhancement comes from. In DRAN, every DU is collocated with its RU at the cell site. Hence, a cell needs independent "housing" facility, DU remains active all the time, and no base-station coordination is deployed at the cell. In traditional CRAN, although DUs are co-located in the DU hotel, there is no sharing of DUs and wavelength, thus every cell needs an active DU and optical transceiver dedicated to service it, and only limited base-station coordination is deployed (ICIC but no JT). We assume simple CRAN architecture as an intermediate state of RAN evolution because it helps us understand where the superiority of V-CRAN comes from.

4.1. Increase in System Throughput

In V-CRAN, VPON and V-BS can be jointly formed with the objective to maximize total throughput of all users, considering the design principles described in Sections 2 and 3.2.

In Fig. 3, we present simulation results to compare throughput performance of V-CRAN, CRAN, and DRAN architectures. In Fig. 3a, we plot average throughputs for cell-average and cell-edge users, respectively, with changing of utilization factor ("u-factor") (for definition, see Fig. 3). DRAN achieves lowest throughput and suffers sharp throughput degradation because there is more interference when u-factor becomes larger. CRAN achieves better performance than DRAN (at most 23% and 573% for cell-average and cell-edge users, respectively), because ICIC can reduce interference, especially for low u-factor, where it is easier to avoid overlapping between RB-user assignments when RB resources are sufficient. V-CRAN can further enhance the throughput because of JT, with about 25% improvement for cell-average users compared to DRAN, and throughput enhancement is more significant for cell-edge users (almost 645%). But when u-factor becomes larger, there is less throughput enhancement for V-CRAN, because a VPON can accommodate fewer cells, and thus less JT services can be provided.

In Fig. 3b, we plot cell-edge throughput of the three architectures for available bandwidth of each cell ranging from 5 to 20 MHz under two scenarios: low and high u-factors. Confirming Fig. 3a, V-CRAN achieves the highest throughput for various spectral bandwidth availability. We note the bandwidth saved by V-CRAN and CRAN to achieve the same throughput of DRAN with 20 MHz configuration. For low u-factor, V-CRAN and CRAN save around 8.5 MHz (43%) and 7.1 MHz (35.5%), respectively. For high u-factor, V-CRAN and CRAN with smallest available bandwidth can achieve higher cell-edge throughput than DRAN with largest bandwidth. V-CRAN's superiority is more noticeable when u-factor is small, confirming Fig. 3a. Results in Fig. 3b show that the evolution of network architecture can not only bring higher data-transmission rate, but also more efficient use of precious spectral bandwidth.

In Table 1, we list the amounts of various resources needed to achieve the cell-edge throughput of DRAN with 20 MHz configuration in the three architectures. DRAN consumes as many DUs as number of cells because of collocation of DU with RU,

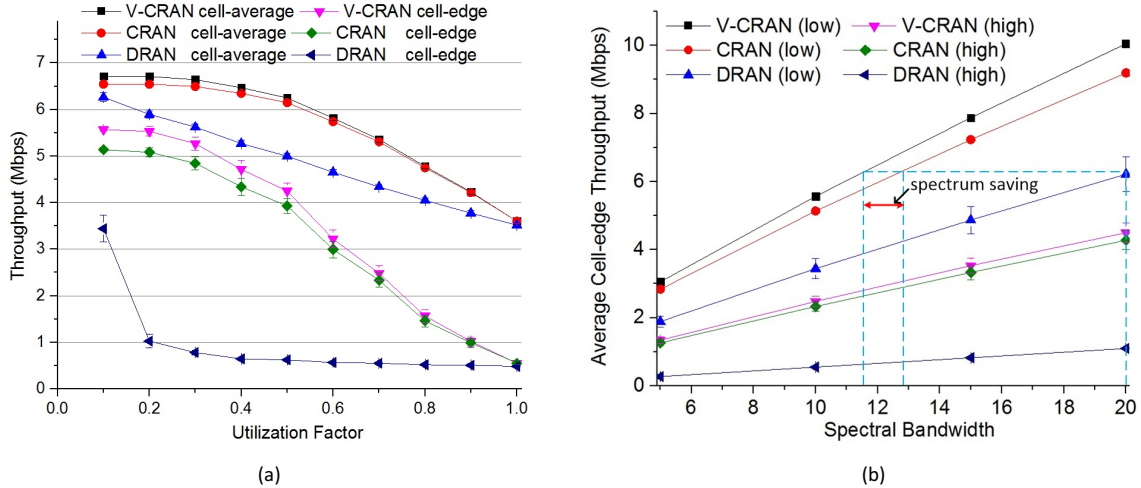


Figure 3. Comparison of throughputs achieved by V-CRAN, CRAN, and DRAN. Simulation experiments are conducted on a 19-cell hexagonal-like cellular network in urban area (with wrap-around and 3GPP urban path-loss model). The inter-cell distance is 500 meters. We assume 10 MHz LTE system for each cell. RBs of a cell can be grouped as larger resource block groups (RBGs), conceptually the same with “resource block” in this article, and assigned to users equally where each user gets maximum one RBG. Users are stationary and uniformly distributed in the network. Utilization factor (u -factor) is the ratio of the number of users (equal to the number of occupied RBGs in DRAN scenario) to the total number of RBGs in the network. Bandwidth of a wavelength is 10 Gbps. a) Comparing the average throughput (per user) of three architectures for all users (cell-average) and 5% worst case users (cell-edge). b) Comparing the cell-edge throughput at low u -factor (0.1) and high u -factor (0.7), respectively. More details can be found in [12].

Table 1. Resource consumption of V-CRAN, CRAN, and DRAN to achieve cell-edge throughput of DRAN with 20 MHz at low utilization factor.

	Spectral Bandwidth	Number of DUs	Number of Wavelengths
V-CRAN	11.5 MHz	2	2
CRAN	12.9 MHz	19	19
DRAN	20 MHz	19	0

but it does not need the optical transport network. CRAN consumes same number of DUs and wavelengths with the number of cells, and it can save some spectral bandwidth. V-CRAN consumes much less spectrum, DUs, and wavelength resources. Note that, although V-CRAN has slight improvement over C-RAN in terms of throughput, it is much more resource efficient, thanks to the virtualization and sharing of VPON and the V-BS concept by utilizing 89% less wavelengths and DUs than C-RAN.

4.2. Enhancement in Energy Efficiency

V-CRAN can save energy because of two aspects: pooling resources and forming cell-centric V-BS. First, pooling DUs can save a large amount of energy consumed by housing facilities, which require power to ensure proper operational conditions (e.g., cooling) of DU, although they do not perform network functions. Although this decoupling of DU and RU needs a transport network, TWDM-PON provides an energy-efficient solution because optical fiber and splitter are passive devices that consume little energy, and OLT and ONUs are low-energy-consuming devices.

Second, energy is saved by forming V-BS for each cell using “just enough” virtualized resources, and timely reclaiming of extra resources with variations of traffic demands. Also, we can shut down more DUs proactively by offloading their remaining traffic load to others through UP redirection described in Section 3.1.

In Fig. 4a, we plot energy consumptions of the three architectures during a day. On a per-day basis, V-CRAN saves 46.1% and 84.1% power consumption compared with CRAN and DRAN, respectively. By comparing CRAN and DRAN, we see the savings in power consumption (and enhancements in energy efficiency) due to pooling of resources, because we only have DU pooling but no sharing mechanism for the reference CRAN architecture. V-CRAN can further save power consumption (and further enhance energy efficiency) compared with CRAN because of V-BS formation. By forming V-BS adaptive to the traffic variation during a day, a large amount of power can be saved by shutting down unused network components during idle hours.

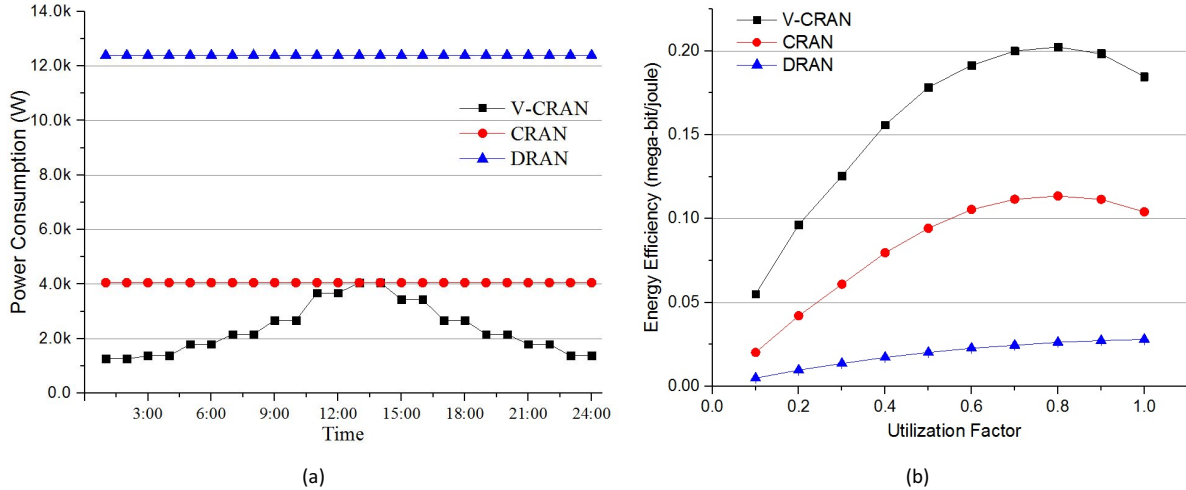


Figure 4. a) Comparing power consumptions of V-CRAN, CRAN, and DRAN architectures during a day. The daily 24 hours are slotted into 12 periods, each of 2-hour length. Users have busy hours from 10:00 to 17:00, and traffic loads reach the peak at 13:00-14:00. b) Comparing the energy efficiency of three architectures for different load ratios. Note that energy efficiency is the number of bits that can sent by consuming one joule. More details can be found in [11].

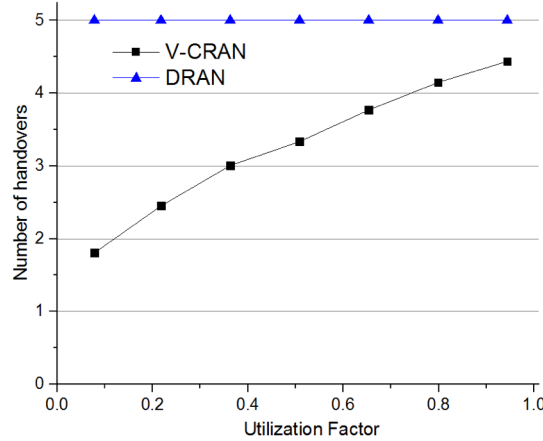


Figure 5. Comparison of average number of handovers suffered by a user for different load ratio when it is traversing the network. Simulation experiments are conducted in a network with one DU-cloud connected with 100 cells by a TWDM-PON. Modified random waypoint mobility model is assumed. Every user gets connected to the network for a time duration, when it is supposed to keep the communication active and must traverse 5 cells before it gets disconnected with the network. Such a user arrives for getting connected to the network following a Poisson distribution. 100000 such users are simulated. This modification is made because we want to fix the number of handovers in DRAN to benchmark the handover reduction achievable for V-CRAN while eliminating the confounding factors (e.g., randomness) induced by the mobility model. Note that we did not plot the performance of CRAN because we assume there is no JT implemented in traditional CRAN architecture, and thus it has the same average number of handovers with DRAN.

In Fig. 4b, we further compare the energy efficiency of three architectures. V-CRAN achieves much higher energy efficiency than other two references because it can enhance the throughput and reduce power consumption. We find that optimal energy efficiency can be achieved when u-factor is around 0.8.

4.3. Improvement in Mobility Management

When mobility patterns of users are considered, number of handovers can be minimized by forming a user-centric V-BS for a mobile user.

In Fig. 5, we plot average number of handovers suffered by a user in V-CRAN and DRAN for different u-factors. V-CRAN needs fewer handovers for various u-factors compared with DRAN. Handover reduction is more noticeable when u-factor is lower, because of two reasons. First, number of inter-VPON handovers is less because, when traffic load in each cell is small, a single VPON can cover a larger area, and DU can provide more base-station coordination for users moving within the area. Second, intra-VPON handover is also less because, when traffic load is low, there are plenty of RBs available in each cell, so there are more

opportunities to find the identical available RB in cooperating cells for a user. But when u-factor increases, both numbers of inter-VPON and intra-VPON handovers increase, because VPON is smaller and RBs are less.

5. Conclusion

5G mobile networks need to provide not only higher data rates but also support diverse quality-of-service requirements coming from emerging mobile services and user equipment. Traditional distributed radio access network (DRAN) is not scalable and cost-efficient for managing the expanding network infrastructure and resources in a flexible manner to adapt to different service requirements. Cloud RAN (CRAN) centralizes the digital unit (DU) of base station (BS) to a central office through a high-speed optical transport network, viz. time-wavelength division multiplexing passive optical network (TWDM-PON). But simply pooling them together cannot fully achieve the gains of CRAN. In this article, we presented the virtualized CRAN (V-CRAN) that virtualizes network resources, including DUs, BSs, and TWDM-PON between them. A joint framework, virtualized base station that can be formed for either a cell or a user, was proposed for resource sharing and base-station coordination. Within this framework, joint optimization of heterogeneous resources can be achieved in 5G transport networks. We presented several use cases of V-CRAN to show how network architecture evolution can enhance system throughput, energy efficiency, and mobility management.

For future work, the impact of different function splits between DUs and RUs on the performance of use cases in V-CRAN can be evaluated. Also, it will be interesting to study heterogeneous environments characterizing 5G, where different types of BSs and radio transmission techniques coexist.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0132-15-1004, SDN based wired and wireless converged optical access networking).

References

- [1] D. Soldani et al., "A 5G Infrastructure for Anything-as-a-Service," *J. of Telecommunications System & Management*, vol. 3, no. 2, 2014.
- [2] Next generation mobile network alliance, "RAN Evolution project CoMP evaluation and enhancement," Mar. 2015.
- [3] China Mobile Research Institute, "C-RAN: The road towards green RAN V3.0," Dec. 2013.
- [4] T. Pfeiffer, "Next Generation Mobile Fronthaul Architectures," *Proc. OFC*, Los Angeles, USA, 2015.
- [5] ITU-T G.989 series of Recommendations, "40-Gigabit-capable passive optical networks (NG-PON2)," Mar. 2013.
- [6] D. Iida et al., "Dynamic TWDM-PON for Mobile Radio Access Networks," *Optics Express*, vol. 21, no. 22, pp. 26209-26218, 2013.
- [7] N. Shibata, et al., "Dynamic IQ Data Compression Using Wireless Resource Allocation for Mobile Front-Haul with TDM-PON [Invited]," *IEEE J. of Optical Communications and Networking*, vol. 7, no. 3, pp. A372-A378, Mar. 2015.
- [8] Next generation mobile network alliance, "Further Study on Critical C-RAN Technologies," Mar. 2015.
- [9] C. L. I, et al., "Rethink fronthaul for soft RAN," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 82-88, Sep. 2015.
- [10] P. Chunyi et al., "Traffic Driven Power Saving in Operational 3G Cellular Networks," *Proc. ACM Mobicom*, USA, 2011.
- [11] X. Wang et al., "Energy-Efficient Virtual Base Station Formation in Optical-Access-Enabled Cloud-RAN," *IEEE J. on Selected Areas in Commun.*, vol. 34, no. 6, Jan. 2016.
- [12] X. Wang et al., "Throughput Enhancement in Optical-Access-Enabled Cloud Radio Access Networks," *UC Davis Technical Report*, Jan. 2016.
- [13] B. Haberland et al., "Radio Base Stations in the Cloud," *Bell Labs Tech. Journal*, vol. 18, no. 1, pp. 129-152, May 2013.
- [14] F. Rossi et al., eds., *Handbook of Constraint Programming*, Elsevier, 2006.
- [15] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2. 3GPP TS 36.300 version 10.5.0 Release 10, Nov. 2011.