Dynamic Provisioning for High Energy Efficiency and Resource Utilization in Cloud RANs

Abolfazl Hajisami, Tuyen X. Tran, and Dario Pompili
Dept. of Electrical and Computer Engineering, Rutgers University–New Brunswick, NJ, USA
E-mails: {hajisamik, tuyen.tran, pompili}@cac.rutgers.edu

***Abstract*—Current Distributed Radio Access Network (D-RAN) architectures, which are characterized by a static configuration and deployment of Base Stations (BSs), have exposed their limitations in handling the temporal and geographical fluctuations of capacity demand as well as the electromagnetic interference caused by the high band reuse, making them inadequate to support the ever-increasing users' data-rate requests. Cloud Radio Access Network (C-RAN) is a new centralized paradigm based on virtualization that has emerged as a promising architecture to address efficiently such fluctuations. C-RAN provides high energy efficiency and resource utilization across Software Defined Wireless Networks (SDWNs). A novel reconfigurable solution based on C-RAN is proposed to adapt dynamically and efficiently to fluctuations in per-user capacity demand. A real-time testbed is used to compare the proposed dynamic provisioning solution against the traditional static approach.**

***Index Terms*—Cloud Radio Access Network, Software Defined Wireless Network, Virtual Base Station, Energy Efficiency.**

I. INTRODUCTION

User demography and the demand for capacity vary depending on the time of the day and week (so-called *tidal effect*). In a traditional Distributed Radio Access Network (D-RAN), spectral and processing resources of each Base Station (BS) are only used by the active users in its cell, causing idle BSs in some areas/times and oversubscribed BSs in others. Moreover, there is no fixed cell size that optimizes the overall coverage and the energy efficiency of a cellular network. This means that the use of small cells is quite efficient in terms of power consumption as well as utilization of processing resources when the capacity demand is high and evenly distributed in space; however, it becomes less so when the data traffic is low and/or uneven due to the static resource provisioning and fixed power consumption [1]. To address these challenges, a new centralized architecture based on Software Defined Wireless Network (SDWN) has emerged.

Cloud Radio Access Network (C-RAN) is a new architecture for cellular networks where the BSs' computational resources are pooled in a central location [2]. C-RAN consists of three main parts: 1) Remote Radio Heads (RRHs) plus antennae, which are located at the remote site and are controlled by Virtual Base Stations (VBSs) housed in centralized processing pools, 2) the BaseBand Unit (BBU) (VBS pool) composed of high-speed programmable processors and real-time virtualization technology to carry out the digital processing tasks, and 3) low-latency high-bandwidth optical fibers, which connect the RRHs to the VBS pool (Fig. 1). The communication functionalities of the VBSs are implemented (in software) on Virtual Machines (VMs) hosted over general-purpose computing servers that are housed in one or more racks of a small cloud datacenter. In a centralized VBS pool, since all the information from the BSs is resident in a common place, BSs can exchange control data at Gbps speed.

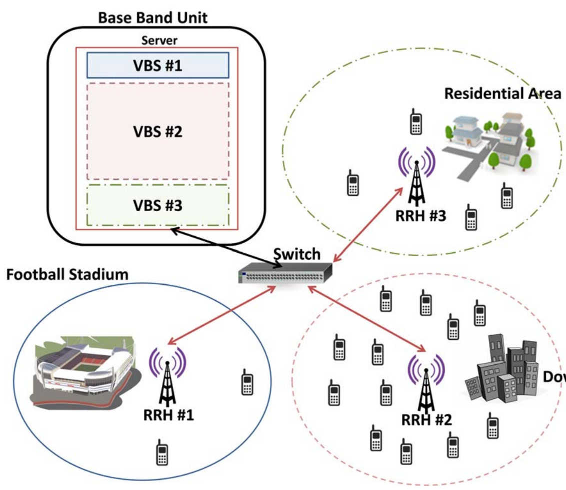


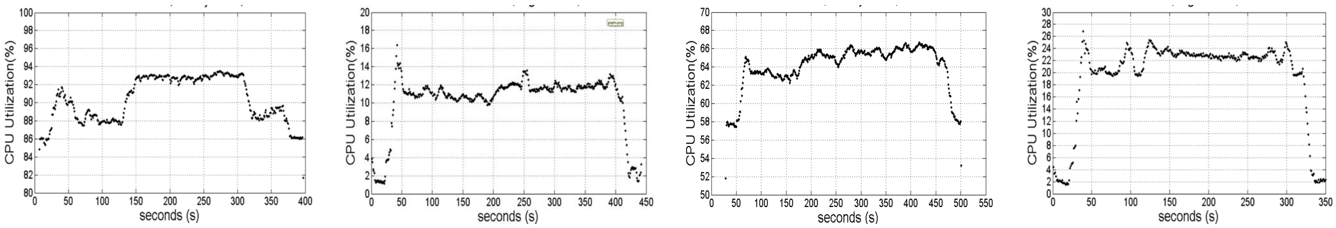
Fig. 1. Cloud Radio Access Network (C-RAN) architecture, where the Base Stations (BSs) are physically unbundled into Virtual Base Stations (VBSs) and Remote Radio Heads (RRHs). The use of virtualization in C-RAN allows for dynamic re-provisioning of spectral and computing resources based on traffic demand fluctuation.

II. PROPOSED SOLUTION

We propose a dynamic provisioning approach aimed at increasing the resource utilization and energy efficiency while providing a high level of Quality of Service (QoS) to the mobile cellular users. We advocate *demand-aware resource provisioning*, in which VBSs will be dynamically 'resized' to meet the fluctuating traffic demands of the cellular network. To do so, we propose to change the VBS size[1] based on traffic demand and user demography. As exemplified in Fig. 1, the VBS with higher user density (cell #2) will be provisioned with more computing resources compared to the ones with lower user density (cells #1 and #3).

Our solution for dynamic provisioning (or re-provisioning) of VBS resources to handle traffic fluctuations has a *proactive*

[1]The size of a Virtual Base Station (VBS) is represented in terms of its (i) processing power [CPU cycles per second], (ii) memory and storage capacity [Bytes], and (iii) network interface speed [bps].



(a) (b) (c) (d)

Fig. 2. Comparison of static provisioning in a D-RAN architecture against dynamic provisioning in C-RAN in terms of CPU utilization: (a) Heavy load and (b) Light load with static provisioning; (c) Heavy load and (d) Light load with dynamic provisioning.

and a *reactive* component; in the former, the fluctuation in per-user capacity demand is predicted and the computational resources are provisioned in advance for a limited horizon. This anticipation is a result of *knowledge of known patterns* (e.g., day and night, weekdays and weekends, holidays, game schedules, etc.) or *predictions* based on advanced time-series analysis of historical traffic traces from immediate as well as distant past. Once estimates of the number and combinations of different types of mobile-data traffic are available, one just has to look up the closest profile and decide on the amount of resources to be provisioned for the VM. Even though the proactive component allows for a smooth transition and greater optimization in ensuing the VM-allocation procedure in terms of energy expenditure and resource utilization, it cannot handle uncertainties. Some of the causes for uncertainties include unanticipated fluctuations in the number of users and per-user capacity demands in emergency scenarios arising out of natural (e.g., hurricanes, tsunamis) or man-made (e.g., industrial accidents, transportation system failures) disasters, unavailability of certain profiles, inaccuracies in the generated profiles, and mismatch between generated profiles and reality due to hardware performance degradation. Conversely, the reactive component monitors/profiles the CPU/memory/network utilization of the VMs and triggers over- or under-provisioning alerts when there is a 'significant' mismatch between the expected resource utilization (based on the profile) and the actual observation.

III. Preliminary Results

To show the benefits of our solution, we built up a real-time testbed and compared the CPU power consumption of our solution under a C-RAN architecture against the traditional approach in D-RAN. The computer we used in the testbed is a Dell workstation with a 6-core Intel Xeon E5-1650 processor with 12 threads and 32GB of RAM. This allows us to simulate a small datacenter with a large amount of resources for running multiple BSs. We also used Ettus B210 boards as Universal Software Radio Peripherals (USRPs) to establish communication between the BSs and the MSs. We considered two scenarios to compare the performance of our solution with the static traditional approach. In the first scenario, two traditional BSs running OpenBTS [3] on different VMs is considered in such a way that one is operated under heavy load while the other under light load. The heavy-load BS handles 6 concurrent calls and 100 text messages, whereas the light-load BS handles only a single call. Each BS is provisioned with 2 threads to process the data. In the second scenario, 2 VBS running OpenBTS on the same virtual machine is considered under the same loads as described in the first scenario. In this case, since the processing resources are shared, we are able to adjust the BS provisioning. We provision the BS under heavy load with 3 threads and the BS under light load with 1 thread. Figure 2 shows the CPU utilization of the two different scenarios and under different loads. For the first scenario, the heavy load and light load have an average CPU utilization of 77% and 10%, respectively, while for the second scenario the average CPU utilizations are 57% and 19%.

TABLE I
Power usage, CPU utilization, and resource utilization for static (D-RAN) and dynamic provisioning (C-RAN).

Provisioning	Type of Load	Power Usage (Watt)	CPU Utilization	Resource Utilization
Static in D-RAN	Heavy Load	70.62 W	77.47%	1.23GB, 2 threads
	Light Load	64.02 W	10.13%	1.26GB, 2 threads
Dynamic in C-RAN	Heavy Load	72.34 W	57.04%	1.13GB, 3 threads
	Light Load		18.58%	1.12GB, 1 threads

Table I shows that dynamic resource provisioning decreases CPU power consumption for the same amount of traffic from 134.64 to 72.34 W. This 42.26% decrease is due to the fact that in the traditional D-RAN each BS needs to have its own dedicated processor while in C-RAN multiple VBSs can share processing resources and exploit multiplexing gain.

IV. Conclusion

We presented a novel reconfigurable solution based on C-RAN – a new centralized paradigm based on virtualization that has emerged as a promising architecture for broadband wireless access – that is aimed at adapting dynamically to fluctuations in per-user capacity demand and at offering higher energy efficiency and resource utilization. Real-time testbed results showed that our proposed dynamic provisioning solution is able to decrease power consumption by more that 40% with respect to the traditional static approach.

References

[1] D. Pompili, A. Hajisami, and H. Viswanathan, "Dynamic provisioning and allocation in Cloud Radio Access Networks (C-RANs)," *Ad Hoc Networks*, vol. 30, pp. 128–143, 2015.
[2] C. M. R. Institute, "C-RAN: The Road Towards Green RAN," in *C-RAN International Workshop*, Oct. 2011.
[3] Range Networks, *OpenBTS Application Suite*. User Manual, 2014.