Computer Communications 128 (2018) 74–79

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

From cloud-based communications to cognition-based communications: A computing perspective

Min Chen[*,a,b], Victor C.M. Leung[*,c]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
[b] Embedded and Pervasive Computing (EPIC) Lab, Huazhong University of Science and Technology, China
[c] Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, Canada

ARTICLE INFO

Keywords:
Cognitive computing
Cognitive communications
Cloud-based communications
Edge computing

ABSTRACT

Traditional cloud-based communications provide powerful cloud computing services. However, simply supporting intensive data processing is not sufficient, especially when capacity is limited and ultra-low latency is required. Thus, it is critical to propose a new Artificial Intelligence (AI)-enabled heterogeneous networks, including various terminal networks, fogs and clouds. Derived from cognitive science and data analytics, cognitive computing can mimic or augment human intelligence. When such cognitive intelligence is integrated with communications, traditional services will be renovated with higher accuracy and lower latency. In this paper, we propose cognition-based communications, which originates from both AI-based intelligent computing and advances in communications. Then, we introduce two applications of cognition-based communications, including user-centric cognitive communications, and cognitive internet of vehicles. Through cognition-based communications, we can better meet users' needs, provide them with a higher Quality of Experience (QoE), and achieve higher energy efficiency.

1. Introduction

Novel information services and applications are expanding globally with the rapid development of wireless communication and networking technologies. Advanced networks and communications can greatly enhance users experience and have made a huge impact in all aspects of people's lifestyles at home, at work, in social exchanges, and economically. Although these advanced techniques have extensively improved users' Quality of Experience (QoE) [1], they are not adequate to meet various requirements such as seamless wide-area coverage, high-capacity hot-spot, low-power massive-connections, low latency high-reliability, and other challenging scenarios. Therefore, it is critical to develop smart wireless communication and networking technologies to support optimized management, dynamic configuration, and fast service composition. Recent year have witnessed that the fusion of computing and communications exhibits a trend to reach such a goal. Cognitive computing, which is derived from cognitive science and data analytics, can mimic or augment human intelligence [2]. In addition, cognitive computing exhibits great potentials to power smart wireless communications, e.g., in self-driving. An intelligent network can be viewed as an existing network integrated with cognitive and co-operative mechanisms to promote performance and achieve intelligence. Under the new service paradigm, there are various technical challenges and problems that need to be addressed to extensively improve the user's QoE, such as complicated decision making for routing, dynamic and context-aware network management, resource optimization, and in-depth knowledge discovery in complex environments. Artificial Intelligence (AI) plays an important role in the ability of cognitive wireless communications to meet many of these technical challenges. Application potentials include complicated decision making, wireless network management, resource optimization, and in-depth knowledge discovery in complex wireless networking environments. Furthermore, wireless communication and network ecosystems must be upgraded with new capabilities, such as the provisioning of personalized and smart Fifth Generation (5G) network services that are assisted by data cognitive intelligence, advanced wireless signal processing based on deep learning, optimized wireless communication physical layer design based on reinforcement learning, adaptive wireless resource management based on cognitive power, etc.

Though various previous works used the similar terminology of "cognitive communications", they focus on the research related with cognitive radio. For example, Green Cognitive Communications in [3], and Cognitive Device-to-Device(D2D) Communications in [4]. In order to avoid the ambiguity, our proposed architecture is named as

* Corresponding authors.
E-mail addresses: minchen2012@hust.edu.cn (M. Chen), vleung@ece.ubc.ca (V.C.M. Leung).

https://doi.org/10.1016/j.comcom.2018.07.010
Received 19 March 2018; Received in revised form 4 June 2018; Accepted 1 July 2018
0140-3664/

"cognition-based communications". In this paper, we first introduce cloud-based communications [6], including Cloud Radio Access Network (C-RAN) [7] and Name Data Network (NDN). Then, we present the architecture and applications of cognition-based communications. In summary, the main contributions of this paper include:

1) We introduce two representative paradigms of cloud-based communications, i.e., computing-centric C-RAN, and cache-enabled NDN.
2) We propose cognition-based communications with a new architecture that includes two layers, i.e., the communication layer and cognition layer. Specifically, the cognition layer consists of two core cognitive engines, i.e., the resource cognitive engine and data cognitive engine.
3) The benefits of applying the proposed cognition-based communications are illustrated by two archetypal wireless networking applications, i.e., user-centric cognitive communications and cognitive internet of vehicles [5].

The remainder of this paper is organized as follows: Section 2 presents the architecture of conventional cloud-based communications, and give its pros and cons. In Section 3, the evolution from cloud-based communications to cognition-based communications is identified from a computer communications point of view. Section 4 concludes this paper.

2. Cloud-based communications

Cloud-based communications include computation-oriented C-RAN and cache-centric NDN. As shown in Fig. 1, C-RANs offload communication-related computing onto the cloud for centralized processing and management. In NDNs, to decrease the delay of content retrieval, contents are cached in various network nodes in a distributed fashion [2].

2.1. Cloud radio-access networks

Radio Access Networks(RANs) are the foundation of today's mobile cellular networks. Traditional RANs built in second to fourth generation mobile core networks are characterized with the following features: (1) Many Base Stations (BS) are deployed with fixed sector antennas. Each BS covers a small area and only handles phone signals in its coverage range. (2) The service capacity is limited by interference and handover losses. (3) The BSs are built with high infrastructure costs. These features have resulted in low spectral efficiency, some air loss, limited Quality of Service (QoS), and high Capital Expenditure (CAPEX) and Operational Expenditure (OPEX). To address these problems, C-RAN was suggested to upgrade or replace current generation RANs in building 5G or future mobile networks. Furthermore, the C-RAN approach is used to reduce the total cost of CAPEX and OPEX, in order to achieve lower energy consumption and higher spectral efficiency. In C-RAN, the bulky antenna towers used in conventional BS are replaced by many small Remote Radio Heads (RRH). These RRHs operate with little power (solar energy is sufficient) and are easily distributed at a high density in populated user areas. The control and processing in the physical BSs are replaced by using Virtual Base Station (VBS) pools housed in a hierarchy of cloud-based switching centers. A balanced traffic load between the RRHs and the VBS pools is enabled by using a high-speed optical transport network and switches with fiber cable and microwave links. The advantages of using C-RAN are summarized in four aspects: (1) A centralized processing resource pool can support between 10 and 1000 cells with high efficiency. (2) Collaborative communications are used in multi-cell joint scheduling and processing, which solves the air loss and handover problems. (3) C-RAN offers real-time services by targeting the open Information Technology(IT) platform, resource consolidation, and flexible multi-standard operation and migration. And (4) a green-energy and clear mobile telecommunication is realized with much less power consumption, lower operating expenses, and a fast system roll out. Many companies are building C-RAN systems including CISCO and Korean Telecommunication.

2.2. Cache-enabled named data networking

To overcome the limited capacity of backhaul links, paradigms called NDN, "Content-Centric Networking" (CCN), and "Information-Centric Networking" (ICN) have been proposed to handle content-dominated Internet traffic for the RANs (front-haul) and the core networks (back-haul). The problem of content caching is generally divided into two steps: content placement and content delivery. Content placement includes determining what content to cache and where to cache it [8]. Content placement also considers how to download the content to the cache node. Content delivery is concerned with how the content is transferred to the requesting users. In general, when the network traffic is low, network resources are cheap and rich (e.g., in the early morning), and content placement should be conducted. Accordingly, when the network traffic is high, network resources are expensive and scarce, and content delivery should be conducted. Consider a heterogeneous RAN, which consists of a Macrocell Base Station (MBS), Small cell Base Stations (SBS), and user terminals. We can summarize four basic caching placement strategies: local caching, D2D caching, Small cell Base Station (SBS) caching, and Macrocell Base Station (MBS) caching. The mobility of users is an important influencing factor for the cache of a wireless access network. By utilizing the mobility of users, the hit rate of content cache can be increased. For caching in the core network, the content can be cached on the routers. The main issues of cache management are content placement and content replacement. For

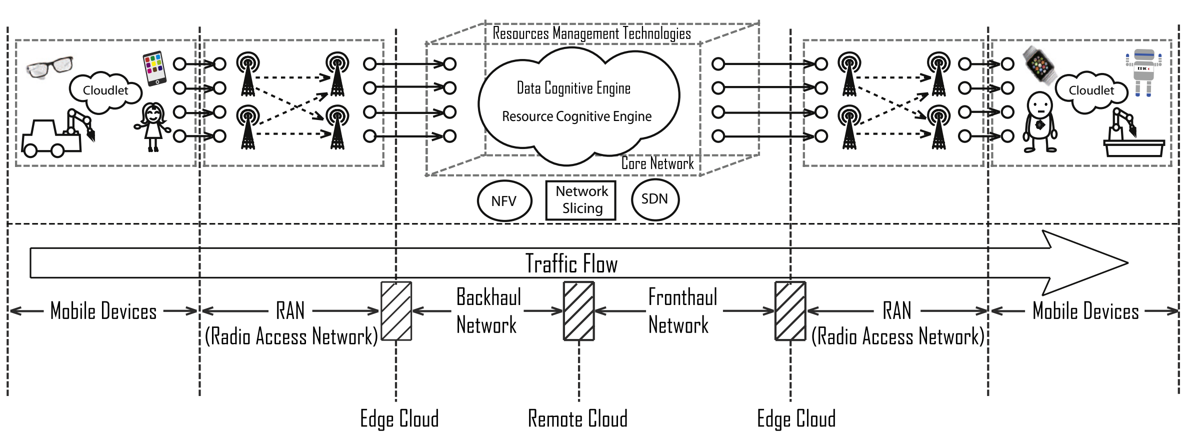

Fig. 1. Cloud-based communications.