

Chapter 14

Online Learning and Heuristic Algorithms for 5G Cloud–RAN Load Balance

Melody Moh

 <https://orcid.org/0000-0002-8313-6645>

San Jose State University, USA

ABSTRACT

The rapidly evolving 5G cellular system adapts cloud computing technology in its radio access networks (RAN), namely Cloud RAN or CRAN. CRAN enables better scalability, flexibility, and performance that allows 5G to provide connectivity for the vast volume of IoT devices. This chapter presents two major research results addressing the load balance (LB) problem in CRAN. First, the authors propose a generic online learning (GOL) system; GOL integrates reinforcement learning (RL) with deep learning method for an environment not fully visible, changing over time, while receiving feedbacks of fluctuating delays. Simulation results show that GOL successfully achieves the LB objectives of reducing both cache-misses and communication load. Next, they study eight practical LB based on real cellular network traffic characteristics provided by Nokia Research. Experiment results of these algorithms on queue-length analysis show that the simple, light-weight queue-based LB is almost as effectively as the much more complex waiting-time-based LB.

INTRODUCTION

Advances in cellular network technology continue to develop to address increasing demands from the growing number of Internet of Things (IoT) devices. A research done by the United Nation's panel on global sustainability stated that, by 2050, 70% of the world population will live in urban areas, which only cover 2% of the entire Earth surface, yet are responsible for 75% of the greenhouse gas emissions (United Nations, 2012). Based on this understanding, the concept of Smart Communities therefore needs the solutions and practices to advance the development and to allow the sustainability of urban environments. In particular, the use of Information and Communication Technologies (ICT) will provide the

DOI: 10.4018/978-1-7998-7708-0.ch014

necessary backbone, not only for maintaining existing services but for enabling new ones; the Internet of Things (IoT) is among the most useful, hopeful ICT technologies for this purpose (Casana & Redondi, 2017). Many expect that the growth of IoT will bring a better quality of life and enable numerous new opportunities.

The rapid growth of IoT also raises new challenges in resource constrained wireless networks, which are the fundamental backbone for IoT services. The emerging 5G communication systems aim to address this challenge by providing ubiquitous, scalable connectivity for vast, exponential growing IoT devices as well as existing cellular applications (Palattela, 2016; Saxena, Roy, Sahu, & Kim, 2017). 5G cellular systems uses emerging technologies such as Millimeter Wave, massive MIMO (Multiple-Input Multiple-Output), and CRAN (Cloud Radio Access networks). These would allow massive connectivity, resource pooling, and energy efficiency needed to support rapid growth of IoT needed for smart urban areas. Among them, the CRAN model is characterized by centralized computation and resource pooling; it therefore promises great scalability and flexibility to support the connectivity of a large volume of devices (Saxena et al., 2017; Checko et al., 2015; Karneyenka, Mohta, & Moh, 2017). More background in CRAN will be described in the background section.

Intelligent systems (IS) have the capability to observe their surrounding environments through their sensory input channels, and interact with the environments via their output channels, through which their actions directly or indirectly affect the environments. Also, a realistic environment is dynamic: with time it changes itself, as well as the elements it depends on.

It is often built to achieve a set of goals. IS in each time-frame requires to make a decision based on the observed data of its surrounding environment. For problems in domains such as robotics, drones, and network controls where the decision (action) is made in real-time, the system is required to adapt in real-time based on the feedback it received. Feedback can be interpreted as some changes in observed data from an environment based on actions of the system. The feedback of a particular action, however, may reach to the system with some delay, or the system may have no knowledge of mapping between observed feedback and its previous actions.

Reinforcement learning (RL) is one of the best learning methods for real-time decision-making (Sutton & Barto, 1998). RL learns from interaction with the environment via recognitions and actions to achieve a goal. On each interaction step, the agent (system), based on the state of the environment chooses an action that alters the state of the environment, and a reward or punishment is then provided to the agent as the desirability of the chosen action. RL algorithms are very useful for solving a wide variety of problems especially when the model is not known in advance; they have been used to solve various problems related to mobile networking and messaging (Muñoz, Barco, Ruiz-Avilés, de la Bandera, & Aguilar, 2013; Shahriari & Moh, 2016; Wu, Lu, Huang, & Zheng 2015).

Deep learning (DL), inspired by the mechanism of human vision, recently attracted more and more attention due to its good performance in many fields such as speech recognition, computer vision, and natural language processing (Hinton & Salakhutdinov, 2006; Krizhevsky, Sutskever, & Sutskever, 2012; Farabet, Couprie, Najman, & LeCun, 2013). As DL has a hierarchical structure, thus it can be applied to any classical learning algorithm.

We proposed GOL (Generic Online Learning), an IS that is able to adapt to dynamic environments without any prior knowledge of the environment or its action policy. GOL forms its abstract definition of state-action space with three hierarchical, deep, layered network consisting of entities. We claim

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/online-learning-and-heuristic-algorithms-for-5g-cloud-ran-load-balance/270197

Related Content

Security and Privacy Challenges in Cognitive Wireless Sensor Networks

Jaydip Sen (2013). *Cognitive Radio Technology Applications for Wireless and Mobile Ad Hoc Networks* (pp. 194-232).

www.irma-international.org/chapter/security-privacy-challenges-cognitive-wireless/78238

A Dynamic Model for Quality of Service Evaluation of Heterogeneous Networks

Farnaz Farid, Seyed Shahrestani and Chun Ruan (2020). *International Journal of Wireless Networks and Broadband Technologies* (pp. 17-42).

www.irma-international.org/article/a-dynamic-model-for-quality-of-service-evaluation-of-heterogeneous-networks/257777

Comprehensive Survey of Routing Protocols for Wireless Body Area Networks (WBANs)

Ali Raza, Pardeep Kumar, Adnan Ahmed and Umair Ali Khan (2020). *Industrial Internet of Things and Cyber-Physical Systems: Transforming the Conventional to Digital* (pp. 145-178).

www.irma-international.org/chapter/comprehensive-survey-of-routing-protocols-for-wireless-body-area-networks-wbans/257844

Link Failure Avoidance Mechanism (LFAM) and Route Availability Check Mechanism (RACM): For Secure and Efficient AODV Routing Protocol

Meeta Singhand Sudeep Kumar (2018). *International Journal of Wireless Networks and Broadband Technologies* (pp. 1-14).

www.irma-international.org/article/link-failure-avoidance-mechanism-lfam-and-route-availability-check-mechanism-racm/209431

Service Discovery in Wireless and Mobile Networks

Hitha Alex, Mohan Kumar and Behrooz A. Shirazi (2005). *Wireless Information Highways* (pp. 251-272).

www.irma-international.org/chapter/service-discovery-wireless-mobile-networks/31450