Expert Systems with Applications 40 (2013) 984–994

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Optimization of load balancing using fuzzy Q-Learning for next generation wireless networks

P. Muñoz *, R. Barco, I. de la Bandera

University of Málaga, Communications Engineering Dept., Spain Campus de Teatinos, Málaga 29071, Spain

ARTICLE INFO

Keywords:
Fuzzy Q-Learning
Fuzzy Logic Controller
Handover
Load balancing

ABSTRACT

Load balancing is considered by the 3GPP as an important issue in Self-Organizing Networks due to its effectiveness to increase network capacity. In next generation wireless networks, load balancing can be easily implemented by tuning handover (HO) margins, achieving a decrease in call blocking. However, call dropping can be increased as a negative effect of the HO-based load balancing, because users usually are handed over to cells where the radio conditions are worse. In this work, a Fuzzy Logic Controller (FLC) optimized by the fuzzy Q-Learning algorithm is proposed for the load balancing problem, with the aim of decreasing call blocking in congested cells, while at the same time restricting call dropping in neighboring cells according to the network policy. In particular, two different approaches for the FLC optimization are evaluated in this work, highlighting that one of the proposed methods allows to accurately preserve the call quality constraint during the load balancing, while the other can adapt to network variations. Results show that the optimized FLC provides a notable reduction in call blocking while preserving call dropping under the operator constraint.

1. Introduction

During the last years, cellular networks have experienced a large increase in size and complexity. Generally, network planning provides proper dimensioning of radio resources during the design phase of the Radio Access Network. However, as traffic demand changes over time, both traffic demand and network resource dimensioning become misaligned, thus leading to an inefficient use of resources. To cope with such a problem in a cost-effective manner, self-optimizing techniques remains as the best solution rather than adding new resources.

Load balancing is a major issue addressed in the field of Self-Organizing Networks (SONs), whose main objective is the reduction of operational effort and cost in networks, (Döttling & Viering, 2009; 3GPP TR 36.902, 2009). SONs consist of a set of principles and concepts that can be applied on advanced real-world networks in order to automate the operation and improve network quality. SONs functions are classified into three related terms (INFSO-ICT-216284 SOCRATES, 2008): Self-Configuration, Self-Optimization and Self-Healing. The first term, Self-Configuration, attempts to automate network deployment and parameter configuration, e.g., when a new resource is added to the existing infrastructure. Self-Healing is related to failure detection, diagnosis and mitigation to cope with major service outages. Lastly, Self-Optimization aims to dynamically adapt network parameters to improve network quality. Load balancing is tackled here as a Self-Optimizing task that can be solved by tuning specific network parameters, like those involved in the handover (HO) process.

The concept of SON is widely addressed in next-generation networks, being part of the specification of the wireless technology known as Long-Term Evolution (LTE) (Rinne & Tirkkonen, 2010). In addition, load balancing, as part of SON, has been recognized as one of the key enablers in LTE-advanced (3GPP TR 36.912, 2011). The load balancing problem can be solved by sharing traffic between adjacent cells. The HO process is responsible for transferring an ongoing call from one cell to another. By adjusting HO parameters settings, the service area of a cell can be modified to send users to neighboring cells. Thus, the size of the congested cell is reduced while adjacent cells increase in size taking users from the congested cell edge. As a result of a better matching between the spatial distribution of traffic demand and network resources, more users could be accepted in the crowded area so that the call blocking probability would be reduced (Luna-Ramírez, Toril, Fernández-Navarro, & Wille, 2011).

A significant research effort has been devoted in the literature to the load balancing problem. Studies are technology-dependent: GERAN (Luna-Ramírez et al., 2011), GSM/UMTS (Tölli & Hakalin, 2002; Pillekeit, Derakhshan, Jugl, & Mitschele-Thiel, 2004), UMTS (Li, Fan, Yang, & Gu, 2005) and LTE (Lobinger, Stefanski, Jansen, & Balan, 2010; Nasri & Altman, 2007; Kwan, Arnott, Paterson,

* Corresponding author. Tel.: +34 952 136 311; fax: +34 952 132 027.
E-mail addresses: pabloml@ic.uma.es (P. Muñoz), rbm@ic.uma.es (R. Barco), ibanderac@ic.uma.es (I. de la Bandera).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.071

Trivisonno, & Kubota, 2010; Wang et al., 2010; Zhang, Qiu, Meng, & Zhang, 2010). Some of the previous references are focused on inter-system load balancing, where propagation conditions are not necessarily a problem because the coverage areas of different radio technologies usually are overlapped. In the case of intra-system load balancing, when a user is handed over to another cell, the propagation conditions could get worse if the user is located near the cell edge because of the interference caused by neighboring cells. In LTE, the previous references analyze the performance of load balancing mostly in terms of achievable throughput for data services, neglecting the dropping rate, which is an important performance indicator in real-time traffic (voice, video, etc.). Also it is noted that controlling those quality indicators (e.g., dropping rate) for real-time traffic is not considered by the previous references as only performance evaluation is carried out without any further impact on the indicators. In Kwan et al. (2010), it is remarked that the trade-off between enhanced blocking and degraded dropping can be adjusted by restricting the maximum achievable HO margin. However, further study of load balancing effects on quality indicators for real-time traffic would be necessary.

Concerning the second topic of the paper, automatic self-tuning implemented by FLCs has been widely applied to mobile network parameter optimization in many references, (Luna-Ramírez, Toril, Ruiz, & Fernández-Navarro, 2008; Werner, Voigt, Khattak, & Fettweis, 2007; ZahirAzami, Yekrangian, & Spencer, 2003). In Toril and Wille (2008), traffic sharing in GERAN is addressed by designing an FLC which iteratively adjusts the HO margins to minimize the call blocking in the network. In Rodríguez, de la Bandera, Muñoz, and Barco (2011), the dynamic load balancing in a realistic urban scenario is tackled by applying an FLC. All these references prove that FLCs are very useful for automatic network parameter optimization. Fuzzy Logic benefits come from their ability to translate human knowledge into a set of basic rules, which represent the mapping of the input to the output in linguistic terms. Such rules are derived from the knowledge and experience of a human expert of the system. In cellular networks, the experience is extracted from network operators, who have to manually adjust many network parameters. However, knowledge usually is not always available and different strategies have been developed to adapt or refine rules, e.g. through learning using neural networks, evolutionary computing or reinforcement learning. Some of these techniques have been successfully applied in wireless network optimization problems, such as in Çalhan and Çeken (2010). Q-Learning is a reinforcement learning method particularly appropriate for learning from interaction, when it is often impractical to obtain representative examples of desired behavior for all the situations in which the controller has to act (Sutton & Barto, 1998). Different works applying Q-Learning in wireless network optimization problems can also be found in the bibliography, (Chen, Chang, & Huang, 2009; Nie & Haykin, 1999; Galindo-Serrano & Giupponi, 2010; El-Alfy & Yao, 2011). The inter-system load balancing problem is addressed in Nasri, Samhat, and Altman (2007) by designing an FLC optimized by the fuzzy extension of Q-Learning algorithm in a UMTS network with WLAN hotspots.

This paper investigates the Self-Optimization of a fuzzy logic controller (FLC) to solve persistent congestion problems in future wireless networks. More specifically, the congestion problem is due to an uneven spatial traffic distribution, e.g., when the center of a city becomes crowded. The optimization process is carried out by the fuzzy Q-Learning algorithm with the goal to reduce the blocking probability for voice services while the call dropping is controlled according to network operator constraints. Results are evaluated in a non-uniformly distributed traffic scenario, where the FLC is optimized to fulfill the call dropping constraint.

Load balancing has an impact on the Grade-of-Service (GoS), which includes call accessibility and maintainability. As a result of a better exploitation of the system capacity, GoS is improved. Although HO-based load balancing is an effective method to share traffic in cellular networks, it may cause negative effects on call maintainability. If the HO margin is decreased, the target cell would increase the probability to be preferred than the serving cell (even if the connection quality is worse), so that some users could be handed over to the target cell. Those users, usually located in the cell edge, will experience worse radio conditions in the target cell as a result of applying such a traffic sharing technique. Thus, negative values of HO margins would increase the risk of dropping. The main contribution of this work is the design of an algorithm to control quality performance for real-time traffic during the load balancing process.

Typically, studies found in the bibliography addressing the load balancing problem aim to provide more capacity to a fixed number of users in the system so that they experience higher instant throughput o lower delay, but accessibility is usually not tackled. In future networks, accessibility will be an important feature as services such as VoIP calls are expected to be widely used. In addition, the connection quality loss experienced by some users when load balancing is carried out and controllability of such an effect have not been properly addressed in the literature. This paper addresses how much the connection quality can be decreased when load balancing is carried out depending on the operator policy. Thus, flexibility and easiness from the network operator perspective is provided by simply adjusting a single parameter. Unlike the design in Nasri et al. (2007), the FLC proposed in this paper balances traffic load in a intra-system LTE scenario where call dropping becomes a key issue. In addition, HO margins of the standard HO algorithm are adjusted in this paper, instead of tuning load thresholds, as proposed in Nasri et al. (2007). In Muñoz, Barco, de la Bandera, Toril, and Luna-Ramírez (2011), the fuzzy Q-Learning algorithm is applied to the intra-system load balancing problem. However, the developed algorithm does not address the problem of the connection quality when load balancing is performed, which is the main objective of this paper. In addition, the analysis in Muñoz et al. (2011) is evaluated in a GSM network. Thus, the FLC proposed in this paper can be considered as a cost-effective solution to increase network capacity in next generation wireless networks as hardware upgrade is not necessary.

The rest of the paper is organized as follows. Section 2 describes the system model for an LTE network, including the main system measurements and HO parameters. In Section 3, the structure of the proposed self-tuning scheme is presented. Section 4 explains the fuzzy Q-Learning algorithm and its application to the proposed FLC. Section 5 describes the simulation setup and discusses the results. Finally, Section 6 draws the conclusions.

2. System model

2.1. Network model

The proposed self-tuning scheme is applicable to any Radio Access Technology for next generation cellular networks. In particular, an LTE downlink macro-cellular network providing constant bit rate service is considered here. Each cell is controlled by an eNodeB (i.e., a base station) and all the cells use the same frequency band. Following the 3rd Generation Partnership Project (3GPP) standard specifications, the eNodeBs are interconnected by the X2 interface, enabling direct communication between them. Thus, measurements such as traffic load can be easily exchanged between eNodeBs over the X2 interface and faster HOs can be performed.