Computer Networks 121 (2017) 112–123

Contents lists available at ScienceDirect

Computer Networks

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

Mobile operators and content providers in next-generation SDN/NFV core networks: Between cooperation and competition

Nir Gazit, Francesco Malandrino*, David Hay

The Hebrew University of Jerusalem, Givat Ram Jerusalem, Israel

ARTICLE INFO

Article history:
Received 31 March 2016
Revised 22 September 2016
Accepted 7 April 2017
Available online 13 April 2017

Keywords:
Software-defined networking
Network function virtualization
Coopetition

ABSTRACT

Much of the traffic processing that today takes place at cloud servers on the Internet will be performed *within* the cellular core network. In such a scenario, different network entities, e.g., switches and servers, will belong to different parties, including traditional mobile operators and over-the-top content providers. The relationship between them is often termed *coopetition*: on the one hand, they are natural competitors; on the other hand, they can reap significant benefits from limited and defined cooperation. Our main purpose in this paper is to model these heterogeneously-owned next-generation networks and solve, within such a novel and complex scenario, the traditional VNF placement and traffic steering problems. Specifically, (i) we present an efficient, online algorithm to make placement and steering decisions, and (ii) evaluate its performance in a realistic scenario. We find that our algorithm would allow mobile operators and content providers to reduce their reliance on third-party vendors – hence the associated costs – by as much as 60%.

1. Introduction

Several evolution trends in cellular networks are now converging and combining with each other.

The first trend concerns the core network and is represented by the emergence of software-defined networking (SDN) and network function virtualization. Today's LTE core networks, based on the Evolved Packet Core (EPC) architecture [1] are built from propriety switches and middleboxes. These include, for example, a Serving Gateway (S-GW) that manages user handovers and billing, a Packet Data Network Gateway (PDN-GW) that handles connectivity to external networks, and cellular endpoints (eNodeB) that provide also encryption and wireless channel management for the network operator. It is important to notice that the EPC architecture already provides a clear distinction between user- and data-plane protocols and protocol entities, however it still uses proprietary and expensive hardware. Thus, a natural direction [2] that next-generation networks might take is to replace these special-purpose boxes with smaller, more flexible middleboxes, each implementing a *network function*.

The second trend mostly concerns the access network, and is represented by *heterogeneity* [3,4]: present-day LTE networks are already composed of different kinds of infrastructure, from macro base stations to femtocells; in the future, LTE-Advanced and 5G networks will integrate multiple radio access technologies, including Wi-Fi [5] and millimeter-wave antennas [6]. In parallel, the traffic served by such networks will also become increasingly heterogeneous, coming from different applications (web browsing, real-time gaming etc.), each requiring different service levels [7] and traversing a different chain of network functions.

The third, and perhaps most disruptive, trend is that, for a variety of reasons, most of which are non-technical [8,9], content providers such as Google, Facebook or Netflix are starting to deploy their own networks [10]. The purpose of such networks is to serve some of the content provider demand directly, and, by that, enhance the available capacity where needed. Unlike the networks deployed by mobile operators, content providers' networks will have a spotty coverage, typically concentrated in densely populated areas. In a similar way, content providers will not build a complete core network, but rather rely on third-party cloud vendors, as depicted in Fig. 1.

However, there is another party that could process the content providers' traffic – mobile operators themselvs. Recall that, in our scenario, mobile operators have built complete core networks capable of processing traffic through VNFs. Furthermore, such networks must be dimensioned so as to meet the *peak* demand mobile operators will face. It follows that, due to daily fluctuations in the traffic demand, they will be underutilized for most of the time.

This spare capacity can be leased to content providers, as shown by the green line in Fig. 1, with significant benefits to both

* Corresponding author:
E-mail address: francesco.malandrino@polito.it (F. Malandrino).

http://dx.doi.org/10.1016/j.comnet.2017.04.014
1389-1286/

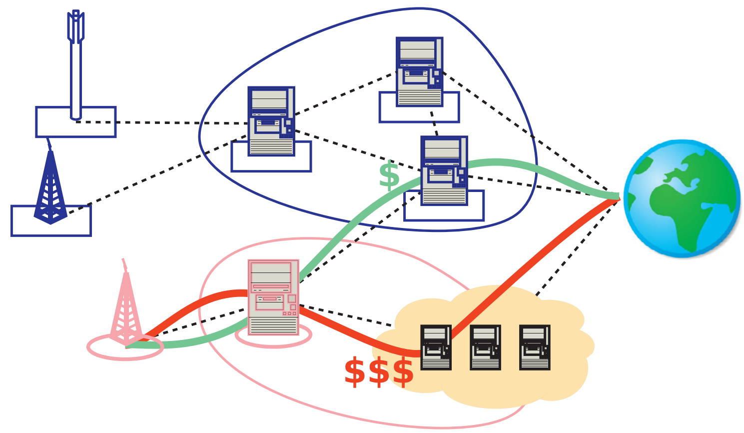

Fig. 1. A next-generation cellular network, composed of both base stations and servers, wherein traffic processing takes place. Both base stations and servers have *owners*: some belong to the blue mobile operator (square shadow), and others to the pink content provider (round shadow). In addition to its own servers, the content provider can have the traffic generated at its own base station (the pink one at the bottom) processed at a third-party cloud vendor (bottom, dark red line) or at servers in the mobile operator's core network, encircled in the blue ellipse (top, light green line), if spare capacity is available therein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parties: mobile operators can obtain further revenue form their core networks, and content providers can save over hefty cloud fees. As the time evolution of different types of demand tends to be different [11], mobile operators are likely to have spare capacity available for content providers exactly when they need it. Furthermore, if no spare capacity is available, content providers can always fall back to cloud providers, as shown by the red line in Fig. 1. In this paper, we will focus on the infrastructure corresponding to VNFs (e.g., commodity servers in the core network), and study how different cooperation level between the mobile operator and content provider yields a different preference of a VNF placement.

Specifically, our contribution is threefold. First, we derive a proper mathematical model for the problem described above. Second, we present an efficient, online algorithm to solve this problem. In online settings, where we need to adapt to changes in demand, our algorithm tries to simultaneously minimize the number of placement changes (thus leading to a small amount of resource-consuming migrations) and maximize the policies satisfied. Finally, we assemble a realistic network trace, using real-life topology and demand information, and verify that the traffic load of content providers and mobile operators is indeed sufficiently different to make a cooperative approach viable. We find that optimal placements can allow mobile operators and content providers to reduce their reliance on third-party providers by 93%. The online algorithm, on the other hand, is able to obtain 60% reductions and scales well for large networks.

It is important to point out that our reference scenario only depicts the simplest possible interaction between mobile operators, content providers, and cloud vendors. More complex interactions, e.g., where prices are dynamically adjusted as the demand evolves, are possible and indeed likely in the real world. By focusing on a simple scenario and on basic interactions, our study is able to capture the essential, qualitative features of heterogeneously-owned networks, while also representing a good starting point for further, more sophisticate works.

At the same time, while next-generation mobile networks are the main motivation and primary reference scenario of our study, the problems we explore – most significantly, VNF placement and traffic steering – are also found in other kinds of networks [12]. Therefore, our solution concept can be successfully applied in any workload placing scenario, where two or more entities with different (and potentially conflicting) objectives have the option to cooperate in order to optimize the overall placement.

We begin by reviewing related work in Section 2, focusing on the evolution of cellular networks and problems akin to ours. Then, in Section 3, we present an optimization formulation of our problem, which we view as a matching between virtual network functions and servers. We further discuss the complexity of such optimization, and find it to be NP-hard. Based on such a result, we turn to a heuristic approach: Section 5 describes the decisions that need to be made, and how our model is used in each of them. Section 6 presents an efficient solution strategy, suitable for real-time usage. We evaluate its performance in Section 8, based on the real-world scenario described in Section 7. Finally, Section 9 concludes the paper.

2. Related work

This work mainly considers the challenges and opportunities of transforming cellular networks into software-defined networks. We discuss recent work that has been conducted in these two important fields of research.

Next-generation cellular networks

The general consensus is that next-generation cellular networks will not be just a faster, higher-capacity version of present-day ones. Indeed, new physical-layer technologies, such as massive MIMO [13], and new types of infrastructure, such as millimiter-wave base stations [6], can support altogether new applications (e.g., proximity [14] and machine-to-machine services). *Heterogeneity* is expected to be a distinctive feature of next-generation cellular networks [3,4], that will also integrate Wi-Fi [5] and device-to-device [15] connectivity.

An important problem in such next-generation networks concerns the ownership and management of such networks. Traditional mobile network operators — typically private companies, with little or no public support — find it increasingly difficult to cope with the capital expenditure needed to update their infrastructure [16]; network sharing agreements between operators represent an early response to this problem [7,17].

At the same time, the availability of high-performance mobile networks is critical for the business model of such *over-the-top* content providers as Google and Netflix. This gave rise to several forms of revenue transfer from content providers to mobile operators. For example, some content providers subsidize (some) of their mobile traffic, as in the Facebook Zero [18] project and a similar Twitter program [19], effectively paying some of their users' bill. More recently, content providers have started to deploy their