Computer Networks 85 (2015) 51–62

Contents lists available at ScienceDirect

Computer Networks

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

Smart routing: Fine-grained stall management of video streams in mobile core networks

Jun He, Wei Song*

Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada

A R T I C L E   I N F O

Article history:
Received 13 October 2014
Revised 12 February 2015
Accepted 24 April 2015
Available online 21 May 2015

Keywords:
Mobile core networks
Video service
User context
Stall management
Optimization model
Smart routing

A B S T R A C T

Video traffic has dominated the global mobile data traffic and creates the fundamental need for continuous enhancement and fast evolution in mobile networks so as to accommodate its unprecedented growth. Despite the surging interests in radio access networks (RANs), the latest technologies on dense and heterogeneous wireless networks are shifting the bottleneck of mobile networks to the core networks. However, managing the stalls of video streams in mobile core networks remains challenging. In an evolving mobile system, the core network needs to (i) determine the data rate for each video streaming request, (ii) distribute the video request among multiple sources, and (iii) route the so-generated peer-to-peer flows. In this paper, we exploit user context and propose an optimized routing scheme (termed as *smart routing*) for stall management in mobile core networks, which adaptively schedules data rates with respect to user context and strategically routes so-scheduled video demands. The proposed smart routing scheme simultaneously addresses the above three aspects by formulating them in a joint optimization problem and solving the formulated problem with a fast algorithm with provable approximation guarantee. Computer simulations validate the efficiency of the proposed scheme.

1. Introduction

1.1. Mobile video traffic and system evolution

The rapid proliferation of smart phones and tablets is profoundly changing the way we behave and consume content, shifting more and more data traffic from fixed networks to mobile networks. Global mobile data traffic is expected to explode by nearly 11-fold between 2013 and 2018, predominantly fueled by video traffic [1]. Up from 53% in 2013, mobile video will represent 69% of global mobile data traffic by 2018. Such substantial growing demands in both traffic volume and speed are out-pacing the ability of current mobile networks [2], *e.g.*, the long term evolution (LTE) systems of the Third Generation Partnership Project (3GPP), which is designed as “interconnecting middle-boxes” between public networks and mobile customers. On the other hand, recent developments in dense and heterogeneous wireless networks are shifting the bottleneck of mobile systems from radio access networks (RANs) to the backhauls, *i.e.*, mobile core networks [3,4].

To accommodate the surging demands, components and protocols of LTE core networks are enhanced or re-factored by integrating transparent video caching and incorporating more efficient request-scheduling algorithms [5–8]. Meanwhile, mobile systems are undergoing fast evolution from the current fourth generation (4G) technologies, *i.e.*, LTE/LTE-Advanced, to the fifth generation (5G) [4] with virtualized networking functions [9–11]. Various video sources, such as in-system caching [7], mobile content distribution networks (MCDNs) [12], cloud service points [10] and improved routers [13], will be deployed within mobile core networks so as

* Corresponding author.
E-mail addresses: jhe2@unb.ca (J. He), wsong@unb.ca (W. Song).

http://dx.doi.org/10.1016/j.comnet.2015.04.012

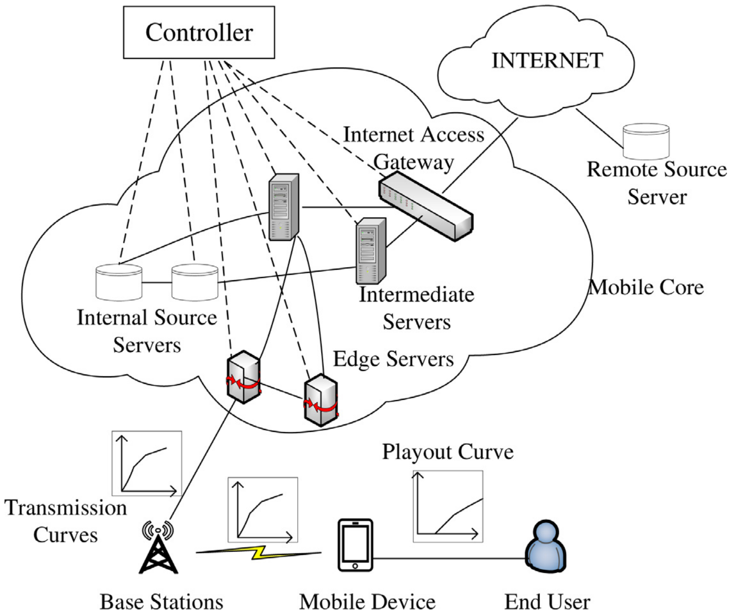

Fig. 1. Video delivery over enhanced/evolving mobile networks.

to move data close to customers. Transport and routing technologies, together with computing and storage resources, will be embedded in mobile core networks to build up a converged infrastructure which orchestrates the delivery of data traffic. To promote manageability and flexibility, the control plane is separated from the data plane in LTE systems and evolving mobile systems. As such, control information is delivered in the control plane while data traffic is transferred in the data plane. This enables centralized networking management in mobile core networks, *e.g.*, through a coordinating component implemented at the packet data network (PDN) gateway of an LTE system or the central controller deployed in a software defined networking (SDN)-enabled mobile core [14].

Considering video services, it is critical and challenging to enhance the quality of experience (QoE) of mobile customers. However, the evolving systems still rely on conventional best-effort routing protocols [15,16] running in lower layers, which fail to make use of the new functionalities to improve QoE and will inevitably let the network bottleneck incline to the core networks. Therefore, mobile core systems are in an urge need of advancing the routing approach with the new network developments and supporting more efficient data delivery and improved QoE.

As we know, QoE is a comprehensive assessment of user-perceived service quality and it depends on a variety of key performance indicators such as latency, data rate, video resolution (spatial and temporal), packet loss and stalling. In particular, stalling is very detrimental to the QoE of video streams [17]. Hence, in this paper, we specifically focus on stall management and aim to provide fine-grained traffic control for video streams in an evolving mobile core network as depicted in Fig. 1. The mobile core consists of a central controller and a set of nodes, including (virtualized) internal content sources,[1] intermediate forwarding servers and edge servers. Edge servers sitting at the edge of the mobile core provide network service to RANs. The solid lines therein represent the data plane, while the dashed lines represent the control plane. In the mobile core network, the central controller periodically gathers network and traffic information, and accordingly shapes network flows and optimizes the network performance.

This architecture is an instant abstraction of future mobile networks with SDN-enabled mobile cores [14]. By mapping the controller to the PDN gateway, edge servers to serving gateways, and intermediate servers to routers, we note that this architecture can also be considered as an enhanced LTE system with built-in content storage. Furthermore, in some cases, internal sources and serving functionalities can be incorporated in evolved routers. This maps the mobile core in Fig. 1 to an information-centric network [13]. Therefore, this model under study is quite general.

1.2. Request routing and stall management

Video streaming requests are initiated by mobile devices and enter the core network through edge servers. Existing works on video streaming support in RANs often assume that the required data are always "pre-fetched" at edge servers [18]. Despite the numerous factors that affect user-perceived service quality, managing the stalls that occur during video sessions is essential to achieve better user experience. As the backhaul is becoming the network bottleneck, efficient request routing in the core network is of vital importance for

[1] These internal sources are virtualized sources using the technologies in [9,10] and can fulfill all the video requests from the controller's view. For example, physical external sources, *e.g.*, CDNs, can be virtualized as internal sources located at the interconnecting gateways.