ARTICLE IN PRESS
JID: INS [m3Gsc;August 10, 2017;13:41]

Information Sciences 000 (2017) 1–21

Contents lists available at ScienceDirect

Information Sciences

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

The cost-efficient deployment of replica servers in virtual content distribution networks for data fusion

Gang Sun[a,b,*], Victor Chang[c], Guanghua Yang[a], Dan Liao[a,d,*]

[a] Key Lab of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China
[b] Center for Cyber Security, University of Electronic Science and Technology of China, Chengdu 611731, China
[c] Xi'an Jiaotong Liverpool University, Suzhou 215123, China
[d] Guangdong Institute of Electronic and Information Engineering, UESTC 523808, China

ARTICLE INFO

Article history:
Received 1 December 2016
Revised 19 July 2017
Accepted 6 August 2017
Available online xxx

Keywords:
Replica severs
Content distribution network
Placement
Network function virtualization
Data fusion

ABSTRACT

Content distribution networks (CDNs) play an important role in distributing content between end users and servers because they can effectively reduce response delays and bandwidth consumption by deploying sets of servers close to end users for service provision. A remaining challenge to CDN deployment is how to efficiently place the replica servers while guaranteeing the quality of service (QoS). As a newly emerging technology, network function virtualization (NFV) facilitates the use of network function software and enables the deployment of different network functions on networked commercial servers. Virtual CDNs are among the most relevant cases that could benefit from NFV. In this paper, we study the problem of efficiently deploying replica servers (middle boxes) in cloud data centers via NFV. We propose an approximation algorithm based on spectral clustering theory to solve the replica server placement problem within an NFV environment. We conduct simulations using local and large-scale data centers to evaluate the performance of our proposed algorithm. The simulation results show that our algorithm achieves lower deployment costs and improved data fusion while optimizing nodes, services and data processing.

1. Introduction

With the development of the Internet, especially wireless network technology, increasing network content has emerged from various applications. The increase in network content has intensified network congestion and lessened users' quality of experience (QoE), which may eventually result in reduced revenue for infrastructure providers (InPs). According to research on online graduate programmes in 2012, 80% of users will abandon connections when the response time exceeds 8 seconds. This behavior can cause financial losses for service providers. Content distribution networks (CDNs) represent an efficient network architecture that can be used to solve the above issue and have been extensively used [43,44]. By placing replica servers as close as possible to the end users, response times can be reduced significantly. Furthermore, core network bandwidth consumption can also be significantly decreased for optimization and to improve data fusion between different services [15,17,34].

* Corresponding authors at: Key Lab of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China.
E-mail addresses: gangsun@uestc.edu.cn (G. Sun), dliao.uestc@gmail.com (D. Liao).

http://dx.doi.org/10.1016/j.ins.2017.08.021

However, the high capital expenditures (CAPEX) and operating expenditures (OPEX) of traditional CDNs obstruct their further development. For example, Akamai built a CDN by deploying more than 6100 servers in approximately 1000 networks in 70 countries [31]. In CDNs, higher service provisioning costs and worse QoE occur as the server-to-user distance increases. Therefore, improving the quality of service (QoS) of CDN and reducing provisional costs remain important issues that must be solved. To improve the performance and lower the CAPEX of CDNs, well-known organizations, including the Internet Engineering Task Force (IETF) and European Telecommunications Standards Institute (ETSI), have pushed standardizations for two different architectures: CDN interconnection (CDNI) [30] and virtual CDN (vCDN) [11]. Both architectures focus on problems in CDNs and are optimal schemes to increase physical resource utilization. However, there are differences between them. CDNI attempts to achieve cooperation among multiple CDN providers to offer better services to end users, whereas vCDN aims to virtualize CDN services in network function virtualization (NFV) environments.

NFV is a promising and critical technology for future network service providers [12] and is expected to change the way services are created, sourced, deployed and supported. In NFV environments, network services are implemented and delivered to users through software only, and they can, therefore, be used to overcome physical constraints and enhance resource utilization. Specifically, NFV supports multi-tenancy and multi-versioning network functions, which allow single physical platforms to be used by different applications, users or tenants [9,28,47].

Motivated by previous work, in this paper, we focus on the problem of efficiently deploying CDN replica servers (middle boxes) in an NFV environment. In our researched CDN replica server placement problem, we sliced the physical resources and abstracted the networked replica servers as a special topology (i.e., vCDN) using NFV technology. Similar to traditional CDN provisioning, we propose an efficient algorithm to decide where to deploy the software replica servers in the abstracted topology (i.e., vCDN). We summarize the main contributions of this paper as follows:

- We present an NFV-based framework for CDN provisioning that abstracts the physical network as a resource topology and then maps the vCDN network function onto the resource topology.
- We model the replica server placement problem as a clustering problem.
- We design an efficient algorithm for solving the problem of optimal replica server placement based on spectral cluster theory.
- By using our steps and approaches, data fusion can be achieved while optimizing nodes, services and data processing.
- We conducted simulation experiments on both known and unknown traffic demands to evaluate the performance of our proposed approach.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 provides the problem statement and formulation. Section 4 presents a detailed description of our proposed algorithm. Section 5 describes the simulation and results of our designed algorithm. Section 6 presents the discussion, and Section 7 concludes the paper.

2. Related work

2.1. Network function virtualization

The goals of NFV are to transform the architecture of traditional networks by using standard virtualization technology to consolidate the network functions of dedicated network equipment onto standard high-volume commercial servers, which may be located in distributed data centers, and to facilitate optimization via the hardware level during the initial phase of service development. As a newly emerging technology for future-generation networks, NFV has attracted substantial attention from researchers [1,2,7,10,27,36,46,48].

For example, in [27], the author presented an NFV approach for innovating the design, management, and operation of a network infrastructure. That article described the progressive evolution of NFV from an initial agnostic software-defined network (SDN) to a fully SDN-enabled NFV solution. The authors in [10] studied the effectiveness of flow features on machine classifiers and the impacts of classifiers for different protocols in an NFV problem. An NFV controller was designed to dynamically make decisions on selecting the most effective classifiers and cost-efficient flow features to use to efficiently categorize network flows. The research presented in [48] explored the problem of optimal provisioning for service chaining according to user dynamical requirements in a network environment. The authors presented a unified control and optimization SDN–NFV framework that jointly controlled and allocated resources in infrastructure networks.

Moreover, in network virtualization, one important issue requiring solution is determining mappings between virtual and physical networks. However, that problem has been proven to be non-deterministic polynomial-time hard (NP-hard) [2] and has thus received substantial research attention. As a result, a large number of approximate or heuristic algorithms have been proposed for solving it. A coordinated node and link mapping approach was presented in [7], in which the authors modelled the optimal virtual network mapping problem through the use of mixed integer programming, relaxed the integer constraints of the model, and then presented two algorithms that they called D-ViNE and R-ViNE. Wang et al. [46] proposed an efficient virtual network mapping algorithm for virtual and physical networks with specific topologies. The research presented in [36] mainly focused on the problem of embedding virtual networks across multiple domains. In [36], the author presented a policy-based framework (i.e., PolyViNE) for inter-domain virtual network embedding. PolyViNE is a distributed protocol that can coordinate the participating InPs and ensure competitive pricing through repetitive bedding