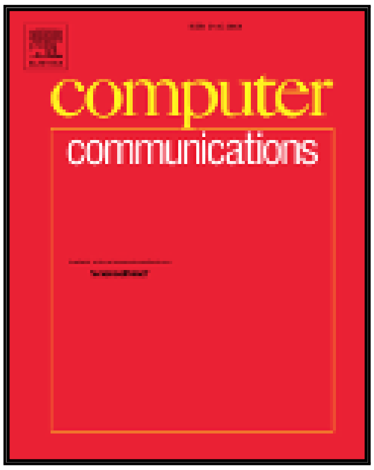
Accepted Manuscript



Joint Load Balancing and Energy Saving Algorithm for Virtual Network Embedding in Infrastructure Providers

Chan Kyu Pyoung, Seung Jun Baek

PII: S0140-3664(17)30701-6
DOI: 10.1016/j.comcom.2018.02.004
Reference: COMCOM 5642

To appear in: *Computer Communications*

Received date: 20 June 2017
Revised date: 29 January 2018
Accepted date: 9 February 2018

Joint Load Balancing and Energy Saving Algorithm for Virtual Network Embedding in Infrastructure Providers

Chan Kyu Pyoung[a], Seung Jun Baek[a,*]

[a]*Department of Computer Science and Engineering, Korea University, Seoul, 02841, Republic of Korea*

Abstract

Network virtualization is key to cloud services, in that it enables multiple users to share a physical infrastructure through abstraction. We propose an online virtual network (VN) embedding scheme which jointly considers load balancing and energy saving so as to maximize the profit of Infrastructure Providers (InPs). For load balancing, we propose to minimize a convex objective which penalizes mapping of VNs to overloaded resources. For energy saving, we consider two popular energy models: speed scaling and power-down. In the speed scaling model, energy consumption is modelled as a convex function of the load imposed on resources. We observe that both load-balancing and energy-saving objectives superadditively penalize high utilization/congestion at resources, and that such synergistic nature of the objectives leads to efficient joint optimization. In the power-down model, a fixed cost exists for keeping a node powered on, which is characterized by a nonconvex energy curve. In this case, we propose an iterative algorithm which explores the trade-offs between load balancing versus cost reduction from power-down of idle servers, in a controlled way. Our algorithm performs a sequential node and link mapping; in particular, for link mapping, we adopt randomized rounding with path stripping in order to obtain a constant factor approximation to the minimum penalty for link utilization. Numerical experiments show the efficacy of our algorithm in servicing VN requests of various topologies and resource requirements.

Keywords: Network virtualization, virtual network embedding, speed scaling, resource allocation, optimization

1. Introduction

The last decades have witnessed tremendous success of the Internet, which ironically has hindered the evolution and innovation of the Internet architecture. Such resistance to change, or *ossification* [1], made it difficult to deploy new technologies and meet diverse service requirements in the existing architecture. Network virtualization [1–4] has gained significant interest so as to overcome the ossification; it allows heterogeneous network services to co-exist in a shared physical infrastructure. Network virtualization is a key enabler for cloud computing services which enjoyed widespread success, e.g., Amazon AWS [5], Google Cloud Platform [6], and Microsoft Azure [7]. Other applications include the virtual testbeds for large-scale network experiments, e.g., PlanetLab [8], GENI [9], Emulab [10].

A set of virtualized end-to-end services are provisioned in the form of virtual networks (VNs). A VN consists of virtual nodes and links accompanied by their resource requirements, where we consider CPU and bandwidth as the node and link resources respectively. The Service Providers (SPs) create VNs by leasing resources from an Infrastructure Provider (InP) in order to provide customized computing services to end-users. The InP overlay the requested VNs onto physical infrastructure network which is called the *substrate network (SN)*. In this paper, we investigate the virtual networking embedding (VNE) problem of how the InP efficiently maps VN requests onto the SN subject to resource capacity constraints. Our goal is to maximize the *profit* of the InP, where the profit in our problem has the meaning of revenue minus operation costs. Thus, we would like to increase the revenue and reduce the cost at the same time. Although the VNE problem is extensively studied recently [11–15], our approach is novel in that it considers the operation profit of infrastructure providers via jointly optimizing revenue and cost. Let us define the revenue and cost in more detail as follows.

1) Revenue: Revenue is defined as the payment received by the InP through leasing resources to SPs. We assume that revenue is an increasing function of the amount of mapped resources. Thus, in order to maximize the revenue, the InP would like to accept as many VN requests as possible, whereas a VN request whose resource requirement exceeds the available capacity is rejected. However, it is known that online VNE problem is NP-hard, e.g., even for exactly determining the feasibility of mapping a VN onto a substrate network without violating node and link capacity constraints at the time of embedding [12]. Moreover, in practice, the InP must make online embedding decisions, which is even challenging due to stochastic nature of arrivals and durations of VNs whose statistics

*Corresponding author

Email addresses: pyoung1101@korea.ac.kr (Chan Kyu Pyoung), sjbaek@korea.ac.kr (Seung Jun Baek)