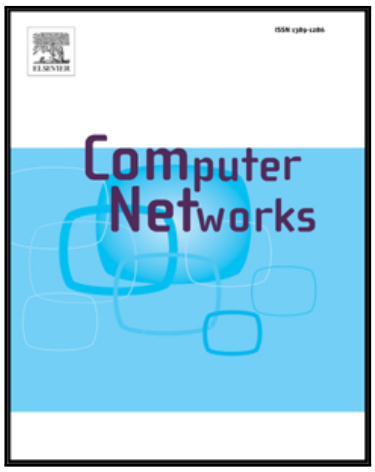


# Throughput Optimization for Admitting NFV-Enabled Requests in Cloud Networks

Zichuan Xu, Weifa Liang, Alex Galis, Yu Ma, Qiufen Xia, Wenzheng Xu

PII: S1389-1286(18)30398-0
DOI: 10.1016/j.comnet.2018.06.015
Reference: COMPNW 6525

To appear in: *Computer Networks*

Received date: 7 February 2018
Revised date: 14 June 2018
Accepted date: 15 June 2018

Please cite this article as: Zichuan Xu, Weifa Liang, Alex Galis, Yu Ma, Qiufen Xia, Wenzheng Xu, Throughput Optimization for Admitting NFV-Enabled Requests in Cloud Networks, *Computer Networks* (2018), doi: 10.1016/j.comnet.2018.06.015

# Throughput Optimization for Admitting NFV-Enabled Requests in Cloud Networks

Zichuan Xu[1], Weifa Liang[2], Alex Galis[3], Yu Ma[2], Qiufen Xia[1,*], Wenzheng Xu[4]

**Abstract**

Network softwarization is emerging as a techno-economic transformation trend that impacts the way that network service providers deliver their network services significantly. As a key ingredient of such a trend, network function virtualization (NFV) is shown to enable elastic and inexpensive network services for next-generation networks, through deploying flexible virtualized network functions (VNFs) running in virtual computing platforms. Different VNFs can be chained together to form different service chains for different network services, to meet various user data routing demands. From the service provider point of view, such services are usually implemented by VNF instances in a cloudlet network consisting of a set of data centers and switches. In this paper we consider provisioning network services in a cloud network for implementing VNF instances of service chains, where the VNF instances in each data center are partitioned into $K$ types with each hosting one type of service chain. We investigate the throughput maximization problem with the aim to admit as many user requests as possible while minimizing the implementation cost of the requests, assuming that limited numbers of instances of each service chain have been instantiated in data centers. We first show the problem is NP-Complete, and propose an optimal algorithm for a special case of the problem when all requests have identical packet rates; otherwise, we devise two approximation algorithms with approximation ratios, depending on whether the packet traffic of each request is splittable. If arrivals of future requests are not known in advance, we study the online throughput maximization problem by proposing an online algorithm with a competitive ratio. We finally conduct experiments to evaluate the performance of the proposed algorithms by simulations. Simulation results show that the performance of the proposed algorithms are promising.

*Keywords:* Throughput maximization; cost minimization; approximation algorithms; online algorithms; network function virtualization; algorithm analysis.

*Corresponding author. Tel.: +8641162274368. Email address: qiufenxia@dlut.edu.cn

*Email addresses:* z.xu@dlut.edu.cn (Zichuan Xu), wliang@cs.anu.edu.au (Weifa Liang), a.galis@ucl.ac.uk (Alex Galis), u5108648@anu.edu.au (Yu Ma), qiufenxia@dlut.edu.cn (Qiufen Xia), wenzheng.xu@scu.edu.cn (Wenzheng Xu)

[1]Dalian University of Technology, P. R. China

[2]the Australian National University, Australia

[3]University College London, UK

[4]Sichuan University, Chengdu, P. R. China