Computer Networks 133 (2018) 1–16

Contents lists available at ScienceDirect

Computer Networks

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

On service-chaining strategies using Virtual Network Functions in operator networks

Abhishek Gupta[a,*], M. Farhan Habib[b], Uttam Mandal[c], Pulak Chowdhury[a], Massimo Tornatore[a,d], Biswanath Mukherjee[a]

[a] University of California, Davis, USA
[b] Intel Corporation, Folsom, USA
[c] Cisco Inc., San Jose, USA
[d] Politecnico di Milano, Italy

ARTICLE INFO

Article history:
Received 9 November 2016
Revised 28 September 2017
Accepted 17 January 2018

Keywords:
Virtual Network Function (VNF)
Network Function Virtualization (NFV)
NFV-capable node
Network-enabled Cloud (NeC)
Service chain
Network-resource consumption

ABSTRACT

Network functions (e.g., firewalls, load balancers, etc.) have been traditionally provided through proprietary hardware appliances. Often, hardware appliances need to be hardwired back-to-back to form a service chain providing chained network functions. These hardware appliances cannot be provisioned on-demand since they are statically embedded in the network topology, making creation, insertion, modification, upgrade, and removal of service chains complex, and also slowing down service innovation. Hence, network operators are starting to deploy Virtual Network Functions (VNFs), which are virtualized over commodity hardware. VNFs can be deployed in Data Centers (DCs) or in Network Function Virtualization (NFV)-capable network elements (nodes) such as routers and switches. NFV-capable nodes and DCs together form a Network-enabled Cloud (NeC) that helps to facilitate the dynamic service chaining required to support today's evolving network traffic and its service demands. In this study, we focus on the VNF service-chain placement and traffic routing problem, and build a model for placing a VNF service chain while minimizing network-resource consumption. Further, we study the network-resource consumption of various service-chaining strategies, which can be a NeC having a varying number of NFV-capable nodes or a proprietary hardware solution. Our results indicate that a NeC having a DC and NFV-capable nodes can significantly reduce network-resource consumption.

1. Introduction

Today's communication networks comprise of a large variety of proprietary hardware appliances (typically called middle-boxes (MBs)) which support network functions such as firewalls, Network Address Translators (NATs), Quality-of-Service (QoS) analyzers, etc. Often, these MBs need to be traversed in sequence, forming a service chain, which provides chained network functions to specific traffic flows. The static allocation of these MBs enforces topological constraints on routing network traffic, as the traffic is required to pass through a set of specific nodes to satisfy service requirements. For example, video traffic requires video-optimization service which is provisioned using two MBs, a video optimizer and a firewall. Hence, video traffic is required to traverse the firewall and then the video optimizer sequentially.

The problem of satisfying service requirements is escalating as, with new cloud applications becoming popular, operators are required to provide more network services to their clients. Current networks support these services with vendor-specific MBs, which are difficult to configure, modify, and upgrade, so the cycle of service introduction, modification, and upgrade/removal is becoming more complex. This problem is compounded because MBs require frequent upgrades due to rapid innovation in technology and increase in traffic volume. This complexity leads operators to route all traffic through the chain, irrespective of service requirements, to avoid misconfiguration and reduce downtime [1].

Network Function Virtualization (NFV) [2,3] provides the operator with the right tools to handle network traffic more effectively and dynamically. As shown in Fig. 1, the predominant idea behind NFV is to replace vendor-specific hardware with Commercial-Off-The-Shelf (COTS) hardware such as servers, switches, and storage [4], which are placed in Data Centers (DCs) or, more generally,

---

[*] Corresponding author.

E-mail addresses: abgupta@ucdavis.edu (A. Gupta), mohammad.f.habib@intel.com (M. Farhan Habib), utmandal@cisco.com (U. Mandal), pchowdhury@ucdavis.edu (P. Chowdhury), mtornatore@ucdavis.edu, tornator@elet.polimi.it (M. Tornatore), bmukherjee@ucdavis.edu (B. Mukherjee).

https://doi.org/10.1016/j.comnet.2018.01.028
1389-1286/

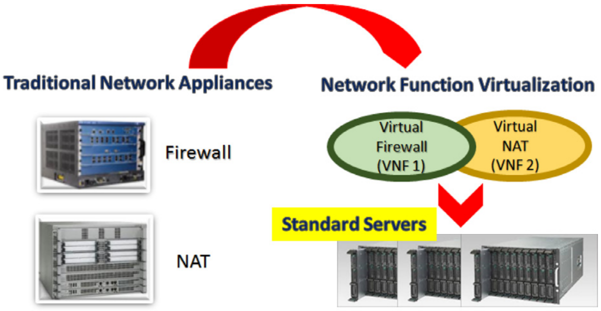

Fig. 1. Network Function Virtualization (NFV) approach.

in network nodes equipped with server capabilities. This leads to more flexibility in deployment of services; and service innovation becomes easier. Operators can scale the service according to traffic intensity while also generating more innovations and revenue. It is for this flexibility in service deployment that VNF-based service chains are being deployed by operators. However, NFV realization has major challenges in key areas of performance, management, security, scalability, resiliency, and reliability [5]. These challenges stem from the necessity of having a “carrier-grade” NFV infrastructure, as network operators are bound by more stringent service-level agreements (SLA) and regulations than a typical enterprise. We refer the readers to John et al. [6] for a summary of the open problems and challenges in NFV.

In this work, we focus on providing a quantitative measure for operators to understand the reductions in Operating Expenditure (OPEX) (i.e., network-resource consumption) to be expected, given the Capital Expenditure (CAPEX) (i.e., number of network nodes capable of hosting VNFs) on NFV platforms. Here, we assume the network-resource consumption to be the main contributor to OPEX because the current DC-based concept of NFV entails frequent redirection of traffic, leading to more bandwidth being consumed. Thus, efficient bandwidth utilization is a major priority for operators.

We develop a mathematical model for sequential traversal of VNFs, i.e., service chaining of VNFs which minimizes the network-resource consumption while ensuring that traffic in the network meets its service requirements. Using our model, we compare the network-resource consumption considering several service-chaining strategies, which can be a NeC having varying number of NFV-capable nodes or a proprietary hardware solution, and we provide a quantitative estimate on the reduction of network-resource consumption achieved when the NFV infrastructure contains network nodes capable of supporting NFV besides the centralized NFV Infrastructure (NFVI), e.g., DCs. This analysis helps an operator to decide which service-chaining strategy is appropriate based on their OPEX targets and CAPEX constraints.

The rest of the study is organized as follows. First, we review relevant literature on VNF placement in Section 2. In Section 3, we elaborate on the concept of VNF, VNF service-chaining and Network-enabled Cloud (NeC). In Section 4, the concept of service chaining is introduced, and various service-chaining strategies are discussed. In Section 5, we describe the VNF service-chaining problem and the model to optimize network-resource consumption. Section 6 presents results for service-chaining strategies. Concluding remarks are provided in Section 7.

2. Related work

The problem of VNF-based service chaining is a combination of traffic routing (multicommodity flow problem) and VNF placement (location problem). A number of studies have appeared recently on both VNF-based service chaining and VNF placement problem. VNFs are run on Virtual Machines (VMs) in a cloud computing environment. So, the problem of VNF placement correlates strongly with VM placement. Several works have investigated the VM placement (VMP) problem. Ref[7]. explains factors that affect resource allocation for VMs in a network-based cloud computing environment. In [8], server selection for VMP is done with respect to optimal routing in a Content Distribution Network (CDN) environment. Su et al. [9] proposes a heuristic for efficient VMP while considering inter-VM relation constraints in a heterogeneous data center scenario. In [10], VMP is solved by considering Service Level Agreement (SLA) using a multi-objective formulation and an evolutionary memetic algorithm. Goudarzi and Pedram [11] solves VMP with SLA using a force-directed search algorithm. In [12], joint multiple resource allocation for VMP is done while accounting for different network delays to a datacenter for different users. Yazir et al. [13] proposes dynamic autonomous resource management in computing clouds using multiple-criteria decision analysis for VMP. In [14], dynamic VMP is done using a multi-cloud application delivery platform called OpenADN. Mills et al. [15] compares multiple VMP algorithms existing in literature. The major distinction between VMP works and VNF service chaining works is the concept of service chaining, i.e., the sequential traversal of VNFs. VNF service chaining needs to account for placement for all VNFs in the service chain such that they can be traversed sequentially by traffic demand.

A number of works focus on the VNF placement problem. Basta et al. [16] looks at the trade-offs of placing LTE mobile core gateways in terms of data-plane delay, control overhead, and number of datacenters utilized using an Integer Linear Program (ILP). Bouet et al. [17] models the problem of placement of Virtual Deep Packet Inspection (vDPI) function as a cost minimization problem which is solved using an ILP and centrality-based greedy algorithm. Similarly, in [18], the authors look at placement of mobile network functions and provide three heuristic approaches for creating and selecting instances of functions. Bhamare et al. [19] minimizes inter-cloud traffic and response time using an ILP and an affinity-based greedy approach when deploying VNFs in a multi-cloud environment. In [20], the objective is to reduce the number of servers deployed for VNFs using an ILP. Cohen et al. [21] provide near-optimal approximation algorithms for the VNF placement problem. Our work differs from the above-mentioned VNF placement literature by accounting for service chaining and formulating it explicitly into our model.

In [22], the VNF placement and routing problem is modeled as a Mixed Integer Linear Program (MILP) to place services optimally for flows and minimize network-resource consumption, and heuristics are developed to place services optimally for a large number of flows but does not do service chaining explicitly. Ghaznavi et al. [23] models the optimal deployment of a service chain as a MILP while optimizing server and bandwidth resources in an intra-datacenter setting. In [24], the authors determine the number of VNFs required and their placement to optimize OPEX while adhering to SLAs using an ILP, while heuristics based on dynamic programming are used to solve larger instances of the problem. Luizelli et al. [25] also models the problem using an ILP to reduce the end-to-end delays and minimize resource over-provisioning while providing a heuristic to do the same. Mohammadkhan co-workers [22,24,25] are limited to an intra-datacenter setting whereas our work applies to a wider operator network. Mehraghdam et al. [26] studies specification and placement of VNF service chains. It develops a heuristic to specify the VNF service chain and a Mixed Integer Quadratically Constrained Program (MIQCP) for the VNF placement problem but does not look at service-chaining strategies required to support NFV. Addis et al. [27] also solves the problem of VNF service-chain placement us-