NFV-enabled Network Slicing

Tachun Lin
Department of Computer Science and Information Systems
Bradley University, Peoria, IL USA
Email: djlin@bradley.edu

Zhili Zhou
United Airlines
Chicago, IL USA
Email: zhili.zhou@united.com

***Abstract*—Next generation (5G) networks are expected to support multiple types of services and various customer segments, which can be realized through the network slicing and network function virtualization (NFV) techniques. Through these techniques, a single physical network can be partitioned into multiple virtual networks providing cost-effective, flexible, and on-demand networking services – the requirements of 5G networks. In this paper, we study the NFV-enabled network slicing problem with limited physical resources. We first present the mathematical formulation to generate individual NFV-enabled network slices as the candidates, followed by the network packing framework addressing users' requirement of multiple network slices. We propose both the mathematical formulation and heuristic algorithm based on column generation as the solution approaches. Our evaluation results show the saving on resource consumptions through network slicing.**

***Index Terms*—Network slicing, network function virtualization, service function chaining, network packing, network service virtualization**

I. Introduction

The next generation (5G) wireless networks target to provide commercial ready telecommunication systems in 2020s [1], which will gather all networks on a platform and support heterogenous services. Network function virtualization (NFV) unleashes NF-related network services from proprietary hardware and enables on-demand NFs execution on commodity hardware through virtualization techniques [2]. Different from network resource slicing in [3], [4], network slicing enabled with NFV in 5G ecosystem constructs network slices which contain dedicated logical networks and their respective networking and computing functions meeting desired KPIs defined by service providers and/or end-users [5], [6]. With network slicing, 5G networks will allow flexible sharing scheme and support dynamic orchestration of services/functions and resources for network operators, service operations, and end-users [7].

A number of design challenges arise when employing network slicing in 5G networks due to the difficulties of orchestration, communication, and transparency among network slices. Leveraging cross-layer network theories, we study in the paper the NFV-enabled network slicing problem from the network design perspective and focus on integrated network slice construction and physical resource allocation supporting multiple network slices in a cost-effective manner. We present a network design problem, "network packing", which allocates physical infrastructure resources to all constructed network slices. We escalate the network covering and packing problem under a single-layer network into the network slicing problem through a cross-layer network setting. Taking advantage of the special problem structure, we then propose a column generation approach, where the master problem deals with physical capacity limitation and the subproblems are for network slice constructions.

II. Related Works

Multiple 5G Infrastructure Public Private Partnership (5G-PPP) projects [6], such as 5G-NORMA [8] and 5GEx [9], are developing, identifying, and testing technologies to support mobile network and radio access network (RAN) network slicing in core networks (CNs), access networks (ANs), and network edges. The optical network also plays an important role in access networks and mobile edges, such as the connection among remote radio heads (RRHs) and baseband units (BBUs) in Cloud-RAN (C-RAN) [10], [11]. Studies showed that virtual networks could be constructed via network embedding and spectrum slot assignment [12], and NFV can be realized [13] in optical networks. In this paper, we take a core network as the physical infrastructure and construct network slices bounded by communication and computational capacities, where each slice not only provides dedicated virtual networks, but also is deployed with required VNFs.

Network virtualization (NV) was considered as slicing physical resources for virtual network construction. Without joint resource management with NFV, network virtualization [14], [15] studied cross-layer network embedding problems to construct single on-demand virtual network. More recent studies on network slicing considered virtual network construction and NF deployment in sequence. [16] demonstrated a RAN slicing architecture, where with virtualized RAN base stations, virtual RAN network is first constructed, and NFs are deployed onto each virtual network system. [17] proposed the logical architecture for mobile network slicing as a two-tier system. [18] presented programmable network slicing architecture with two-level MAC scheduler. The network slicing problem studied in this paper, an integrated NV (network embedding) and NF realization problem, takes network slice construction as subproblems while considering the resource allocation/limitation in the physical infrastructure, simultaneously.

III. Problem Settings and Descriptions

In this section, we first illustrate an instance of single network slice and its related network design problem in Fig. 1, under simple setting, where only non-chained NF services are

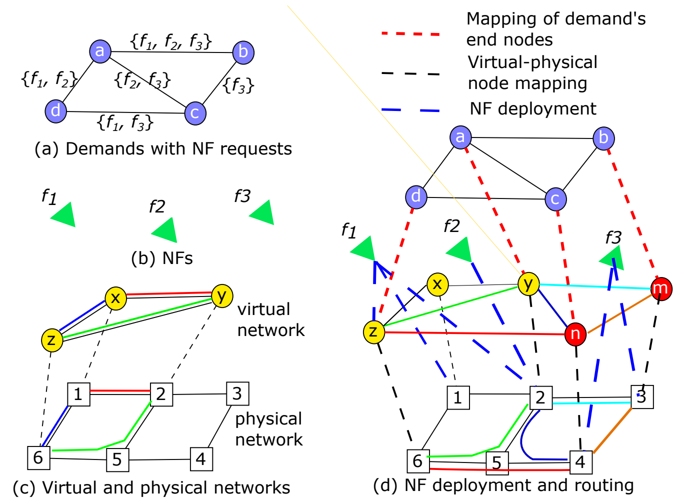

Fig. 1. NFVI design and management

Notation	Description
$G_P(V_P, E_P)$	Physical infrastructure network with node set V_P and edge set E_P
$G_O(V_O, E_O)$	Virtual infrastructure network with V_O and E_O as its node and link sets
Γ	The set of network slices with γ as index
i, j, u, v	Physical nodes, $i, j \in V_P$; virtual nodes, $u, v \in V_O$
e, ϖ	Physical link, $e \in E_P$; virtual link, $\varpi \in E_O$
$m(\cdot)$	Mapping function onto node and path in G_P, respectively
$\psi(\cdot, \cdot)$	Latency evaluation function between two nodes, with $\bar{\psi}$ as the limitation
$\mathcal{F}$	Network function set, with network function $f \in \mathcal{F}$
D^γ	A set of end-to-end demand denoted by a tuple $[(s,t), \sigma_{st}, \Lambda_{st}, \Phi_{st}]$, where s, t are the two end nodes of a demand d_{st}, σ_{st} indicates whether the request is with SFC, Λ_{st} is NF set requested, Φ_{st} is set for required VNF instances
$Q(D^\gamma)$, $N(D^\gamma)$	Node pair set and node set of D^γ
P_{st}, $\overrightarrow{P}_{st}$, $\overrightarrow{P}_{ts}$	Undirected and directed physical path sets with p and η as elements correspondingly

TABLE I
NOTATIONS

considered. Given on-demand virtual requests with required NFs (Fig. 1(a)), all requested NFs (non-chained) are pre-determined (Fig. 1(b)), physical and available virtual resources and their corresponding virtual-to-physical node mapping and virtual edge to physical path routing mapping (Fig. 1(c)). End-to-end demands (demands in brief) are realized through physical infrastructure directly or via available virtual resources. For instance, virtual demand (a, d) is mapped onto virtual link (y, z) and realized through its embedded physical routes $\{(2,5),(5,6)\}$, where NFs f_1, f_2 are deployed onto physical nodes 2 and 6 along the route. Since no virtual node is available for demand (b, c) (violated latency requirements), we augment the virtual network with nodes m, n and links $(y,m),(y,n),(m,n),(z,n)$ and generate their routings on physical network. Note here that the augmented virtual network provides connectivity/routing information for all demands and will be used to fulfill the required network services. This instance indicates that designing on-demand network slicing may require VN augmentation when available virtual resource is not sufficient; and to fulfill NF requests, the VN link mapping may be altered.

Given physical and virtual available infrastructure $G_P = (V_P, E_P)$ and $G_O = (V_O, E_O)$ with C_i and C_e as physical node and edge capacities, respectively, with $i \in V_P$ and $e \in E_P$. We assume that G_O is an abstraction of G_P. G_O's node and edge mappings onto G_P (logical-to-physical node mapping) are known, whose mapping function is denoted as $m(\cdot)$, such that $m(s) = i$ and $m(\varpi) = p_\varpi$ with $s \in V_O$, $\varpi \in E_O$, $i \in V_P$, and $p_\varpi \subset G_P$ (logical edge's routing through a physical path p_ϖ). We let Γ be a network slice set and D_γ be requests of network slice $\gamma \in \Gamma$, whose elements are NF requests represented by a tuple $d^\gamma_{st} = [(s,t), \sigma_{st}, \Lambda_{st}, \Phi_{st}]$ with (s,t) as source and sink nodes of a request, σ_{st} indicates whether the request is with service function chaining (SFC, $\sigma_{st} = 1$) and non-chained NFs ($\sigma_{st} = 0$), Λ_{st} as a list of requested NFs (the order of SFC follows order of NFs in the list), and Φ_{st} as required VNF instances for corresponding NFs. We let $Q(D^\gamma) = \{(s,t) : (s,t) \in d^\gamma_{st}, d^\gamma_{st} \in D^\gamma\}$ and $V(D^\gamma) = \{s, t : (s,t) \in Q(D^\gamma)\}$ represent all node pairs and node set of D^γ, respectively. All notations and parameters are listed in Table I.

Definition 1: Given G_P, G_O, and a network slice γ and its requests D^γ, **virtual network augmentation** is that
(1) virtual node mapping and augmentation: the latency between s and $m(s)$ is within the latency limitation, i.e., $\psi(s, m(s)) \leq \bar{\psi}$; s is mapped into a virtual node q or augmented as virtual node and embedded into physical node i, i.e., $m(s) = q$, or $m(s) = i$ with $s \in V(D^\gamma)$, $i \in V_P$, $q \in V_O$; $|m(s)| = 1$; and $m(s) \neq m(t)$ when $s \neq t$ with $s, t \in Q(D^\gamma)$;
(2) virtual demand pair mapping and embedding: given $(s,t) \in Q(D^\gamma)$, (s,t) maps into available virtual links and utilizes its physical route, i.e., $m(s,t) = m(u,v) = p_{u,v}$; or embedded into physical path, i.e., $m(s,t) = p_{m(s)m(t)}$; and
(3) physical network capacity limitation: all realization are bounded by physical node and link capacities.
Considering the NFV-enabled network slicing, the virtual request realization not only fulfills VN construction conditions, but also takes account of NF service realization. We consider SFCs as logical directed paths, linking all required VNFs in order. To fulfill SFC request, we introduce direction to virtual link mapping and let $\overrightarrow{P}_{st}$ and $\overrightarrow{P}_{ts}$ be the directed paths from s to t and from t to s. We next discuss **SFC request satisfaction** integrated with VN construction as follows:
(1) NF instance deployment: determining NF f instances deployment n^f_i on physical node i with $f \in F$ and $i \in V_P$;
(2) SFC request embedding: (2.1) SFC request mapping and embedding is with a directed physical path in direction $\overrightarrow{m}(m(s), m(t))$ or $\overrightarrow{m}(m(t), m(s))$; (2.2) the embedded directed physical path visiting NF deployed nodes following NF orders in SFC, i.e., $S = \{(i_1, f_1), (i_2, f_2), \cdots, (i_\ell, f_\ell) : (f_1, f_2, \cdots, f_\ell) \in$

$\Lambda_{st}, i_1, i_2, \cdots, i_\ell \in m(m(s), m(t))$ or $m(m(t), m(s))\}$, which records visited physical nodes with corresponding deployed network functions on physical path following path direction. For requests with non-chained NFs, the visiting order of NF deployed nodes in above condition (2.2) could be relaxed.

A **single network slice design problem** is that with given available physical and virtual infrastructure $G_P(V_P, E_P)$ and $G_O(V_O, E_O)$, and virtual request D^γ with $\gamma \in \Gamma$, augmenting virtual network where all virtual NF/SFC requests are satisfied through reaching deployed NF instances via virtual request embedded physical paths. The corresponding optimal single network slice design is with objective minimizing total design costs.

In general, the **network slicing supporting multiple network slices** is that each network slice fulfills single network slice construction conditions; and all together are bounded by physical resource limitations. We introduce network packing problem as follows which manages physical resources for multiple network slices. We let binary variable x_k^γ present a candidate for network slice $\gamma \in \Gamma$, K^γ be candidate set of slice γ, and $\alpha_e^{\gamma k}$ and $\beta_i^{\gamma k}$ present physical link and node resource consumption by network slice γ's candidate k.

Definition 2: Given $G_P(V_P, E_P)$ and $G_O(V_O, E_O)$, and Γ network slices, the **network packing** problem is that
(1) constructing candidates for network slice γ with $\gamma \in \Gamma$;
(2) all network slices are bounded by physical link capacities, $\sum_{\gamma\in\Gamma}\sum_{k\in K^\gamma} \alpha_e^{\gamma k} x_k^\gamma \leq C_e$ with $e \in E_P$; and
(3) bounded by physical node capacities, i.e., $\sum_{\gamma\in\Gamma}\sum_{k\in K^\gamma} \beta_i^{\gamma k} x_k^\gamma \leq C_i$ with $i \in V_P$.
The *maximal network packing* problem is to identify the maximal number of network slices that can be hosts with given physical and virtual infrastructures.

Property 1: No augmented virtual nodes are shared by multiple virtual networks.
Since the network slicing supports multiple virtual networks on a common physical network, on-demand VN construction builts virtual networks to serve NF requests. Hence, no virtual network shares augmented virtual nodes based on virtual requests.

Theorem 1: The single network slice construction problem is NP-hard.
The NP-hard network virtualization problem [15] is a subproblem of the single network slice construction problem.

Theorem 2: The network packing problem is NP-hard.
Since single network slice construction is a subproblem of network packing, the network packing problem is NP-hard.

IV. Solution Approach

In this section, we first present complete mathematical formulation to generate network slice candidates, followed by the solution approach for network slicing via the network packing problem, where x_k^γ represents a candidate for network slicing γ. Table II lists parameters and variables used aftermath.

Parameter	Description
C_i, C_e	Physical node and link capacities
c_i^f	The costs to deploy NF f onto physical node i with $i \in V_P$
c_i, c_e	Costs for network flow routed through node i and link e, respectively, where $e \in E_P$ and $i \in V_P$
$\delta_\eta^i, \delta_p^i$	Binary indicator whether physical node i is on directed physical path η and undirected physical path p
ζ_η^i	The order of node i on directed physical paths η starting from the source node
$\vartheta_{\Lambda_{st}}^f$	The order of network function f in requested SFC Λ_{st}
M	A big number
$\alpha_i^{\gamma k}, \beta_e^{\gamma k}$	Physical node and link resource consumption by slice γ's candidate k with $k \in K_\gamma$, $\gamma \in \Gamma$
c_k^γ	Design cost of slice γ's candidate k with $k \in K_\gamma$, $\gamma \in \Gamma$
Variable	**Description**
h_q^s	Binary variable indicating whether virtual request end node s maps or embeds onto node q where $s \in N(D^\gamma)$ and $q \in V_P \cup V_O$
x_k^γ	Binary variable indicates whether network slice γ's candidate k is selected, if yes, $x_k^\gamma = 1$; otherwise, $x_k^\gamma = 0$
w_γ	Slack variable indicates whether network slice γ cannot be selected due with physical resource limitation, if yes, $w_\gamma = 1$; otherwise, $w_\gamma = 0$
y_η	Binary variable indicates whether directed path η is chosen, if yes, $y_\eta = 1$; otherwise, $y_\eta = 0$ with $\eta \in \cup_{(s,t)\in Q(D^\gamma)} \overrightarrow{P}_{st} \cup \overrightarrow{P}_{ts}$
$\varsigma_{st}^{fi}, \varrho_{st}^{fi}$	Binary variable indicates whether NF f deployed at physical node i for SFC and non-chained request (s,t)
n_i^f	Integer variables representing the number of instances of VNF $f \in \mathcal{F}$ deployed at physical node with $i \in V_P$

TABLE II
PARAMETERS AND VARIABLES

A. Single NFV-enabled Network Slice Construction

In this section, we present an MILP formulation for the single network slice construction, which provides optimal single network slice construction with minimizing design cost as the objective and generates network slice candidates which satisfy all conditions. We consider a network slice $\gamma \in \Gamma$ and present MILP formulation (1SLICE).

$$\min \sum_{f\in F}\sum_{i\in V_P} c_i^f n_i^f + \sum_{\substack{(s,t)\in Q(D^\gamma)\\ \sigma_{st}=1}} \sum_{\eta\in \overrightarrow{P}_{st}\cup\overrightarrow{P}_{ts}} \lambda_{st}\left(\sum_{e\in E_P} c_e\delta_\eta^e y_\eta + \sum_{i\in V_P} c_i \delta_\eta^i y_\eta\right) + \sum_{\substack{(s,t)\in Q(D^\gamma)\\ \sigma_{st}=0}} \sum_{p\in P_{st}} \lambda_{st}\left(\sum_{e\in E_P} c_e\delta_p^e y_p + \sum_{i\in V_P} c_i \delta_p^i y_p\right)$$

s.t.,

$$\sum_{q\in\{j:j\in V_P:\psi(j,s)\leq\bar\psi\}\cup\{j:j\in V_O:\psi(j,s)\leq\bar\psi\}} h_q^s = 1,\ s \in V(D^\gamma) \quad (1)$$

$$\sum_{s\in\{j\in V(D^\gamma):\psi(q,j)\leq\bar\psi\}} h_q^s = 1,\ q \in V_P \cup V_O \quad (2)$$

$$\sum_{\eta \in \overrightarrow{P}_{st} \cup \overrightarrow{P}_{ts}} y_\eta = 1, \qquad \sigma_{st} = 1, (s,t) \in Q(D^\gamma) \tag{3}$$

$$\sum_{p \in P_{st}} y_p = 1, \qquad \sigma_{st} = 0, (s,t) \in Q(D^\gamma) \tag{4}$$

$$(y_{\eta_{uv}} + y_{\eta_{vu}}) - y_\eta \leq 2 - (h_u^s + h_v^t), \quad \sigma_{st} = 1, (s,t) \in Q(D^\gamma), (u,v) \in E_O \tag{5}$$

$$y_{\eta_{uv}} + y_{\eta_{vu}} = 1, \qquad (u,v) \in E_O \tag{6}$$

$$y_{p_{uv}} - y_p \leq 1 - (h_u^s + h_v^t - 1), \quad \sigma_{st} = 0, (s,t) \in Q(D^\gamma), u, v \in V_O \tag{7}$$

$$M(\zeta_\eta^i - \zeta_\eta^j + 1) y_\eta \geq \vartheta_{\Lambda_{st}}^{f_\ell} \varsigma_{st}^{f_\ell i} - \vartheta_{\Lambda_{st}}^{f_{\ell-1}} \varsigma_{st}^{f_{\ell-1} j}, \quad i, j \in V_P, f_\ell, f_{\ell-1} \in \Lambda_{st}, \eta \in \overrightarrow{P}_{st} \cup \overrightarrow{P}_{ts} \quad 1 \leq \ell \leq |\Lambda_{st}|, \sigma_{st} = 1, (s,t) \in Q(D^\gamma) \tag{8}$$

$$\delta_p^i y_p \leq \varrho_{st}^{fi}, \quad \sigma_{st} = 0, f \in \Lambda_{st}, p \in P_{st}, (s,t) \in Q(D^\gamma), i \in V_P \tag{9}$$

$$n_i^f = \sum_{(s,t) \in Q(D^\gamma)} [\varsigma_{st}^{fi} + \varrho_{st}^{fi}] \Phi_{st}^f, \ f \in F, i \in V_P \tag{10}$$

$$h_q^s, y_\eta, y_p, \varrho_{st}^{fi}, \varsigma_{st}^{fi} \in \{0,1\}, \quad s \in V(D^\gamma), i \in V_P, q \in V_P \cup V_O, \eta \in \overrightarrow{P}_{st} \cup \overrightarrow{P}_{ts}, \quad (s,t) \in Q(D^\gamma) \tag{11}$$

$$n_i^f \in Z^+, \qquad f \in F, i \in V_P \tag{12}$$

Constraint (1) determines virtual node s mapped/embedded nodes. Constraints (3) and (4) determine a single directed path and undirected path for SFC request for split NFs, respectively. Constraints (5) and (6) force virtual request to take virtual link embedded physical path, when virtual request mapped onto virtual link. For non-chained NF request, constraint (7) forces single undirected path selection. Constraint (8) forces that virtual request's physical path routes through NF deployed physical nodes following NF order in the requested SFC. Note here that with container based NFV, multiple NFs may deployed at the same physical nodes. Hence, not exact $|\Lambda_{st}|$ physical nodes are utilized. Constraint (8) restricted that no later order NF in a SFC deployed at physical nodes before its previous NF deployed physical nodes. Given two consecutive NFs in a SFC $f_{\ell-1}$ and f_ℓ, their order information is provided by parameter ϑ_{st}^{fi}. The visiting of NF deployed physical nodes following the NF order in the SFC, which requires the visiting of f_ℓ deployed node should not be before $f_{\ell-1}$ deployed physical nodes on the SFC request realized directed path. Constraint (9) indicates whether NF f deployed at physical node i provides NF instances for non-chained (s,t) request. Constraint (10) cumulates physical resources deployed on node i for NF f. Constraints (11) and (12) provide feasible regions for all variables.

The above formulation serves two functionailities: (1) providing the optimal single NFV-enable network slice construction with the minimal construction cost as objective; (2) generating slice candidates. Instead of solving the single network slice construction problem into the optimal, we solve the problem for feasible solutions and utilize feasible solutions to construct x_k^γ.

B. Network Packing Problem

We present mathematical formulation for network packing problem, which determines network slicing supporting multiple network slices with network resource limitations (physical node and link capacities). With a given network slice candidate, its operation costs c_k^γ, and node and link capacity consumptions $\alpha_i^{\gamma k}$ and $\beta_e^{\gamma k}$ could be calibrated. The network packing problem supporting multiple network slices is formulated (NETPACK) as follows:

$$\min_x \sum_{\gamma \in \Gamma} \sum_{k \in K^\gamma} c_k^\gamma x_k^\gamma + \sum_{\gamma \in \Gamma} M w_\gamma \tag{13}$$

$$s.t. \sum_{k \in K_\gamma} x_k^\gamma + w_\gamma = 1, \qquad \gamma \in \Gamma \tag{14}$$

$$\sum_{\gamma \in \Gamma} \sum_{k \in K_\gamma} \alpha_i^{\gamma k} x_k^\gamma \leq C_i, \qquad i \in V_P \tag{15}$$

$$\sum_{\gamma \in \Gamma} \sum_{k \in K_\gamma} \beta_e^{\gamma k} x_k^\gamma \leq C_e, \qquad e \in E_P \tag{16}$$

$$x_k^\gamma, w_\gamma \in \{0,1\}, \qquad k \in K_\gamma, \gamma \in \Gamma \tag{17}$$

The formulation is with the objective to minimize design cost through penalization if network slices with requests cannot be fulfilled due to physical resource limitations. Constraint (14) selects only one candidate for a network slice, or the network slice cannot be constructed due with physical resource limitation. Constraints (15) and (16) provide physical node and link capacity limitations on all network slices. To solve the network packing problem, we present the single NFV-enabled network slicing construction in the next section, with which network slice candidates could be constructed.

C. Solution Approach: Heuristic-based Column Generation

Based on above NETPACK and 1SLICE formulations, we present a heuristic-based column generation [19] [20] [21], where each column is corresponding to single slice candidate which is obtained by randomly generating the 1SLICE cost function.

Algorithm 1 Heuristic-based column generation algorithm for network slicing

Input: $G_P = (V_P, E_L)$, $G_O = (V_O, E_O)$, D^γ with $\gamma \in \Gamma$
for all $\gamma \in \Gamma$, $k = 1, \cdots, \Theta_\gamma$ **do**
  Apply 1SLICE formulation with random deployed cost with deployed costs as mean and varied by random percentage in [-50%, 50%]
  Record slice candidate for γ into K_γ
  Calculate slice candidate design cost, physical node and link resource consumption
end for
Apply NETPACK formulation with all network slice candidates and determine supported network slices

Algorithm 1 first generates network slice candidates via 1SLICE formulation by randomly generating design cost on physical nodes and links, which is a heuristic to generate columns for NETPACK formulation. The number of candidates is determined by parameter Θ_γ as the input. Then, the NETPACK formulation is solved with all generated network slice candidates.

D. Extensions: Network Packing with Multiple Network Services

When other types of network services, such as the content delivery networks, are considered and should be contained in network slices, these network services can be considered as new types of "NF services" which require corresponding dedicated and isolated resource allocation. Hence, by adding one more dimension indicating network service type and corresponding limitations through capacity constraints, our proposed solution approach could support network slicing for multiple network slices and multiple network service types.

V. Experiment Results

In this section, we present our experimental design and simulation results for the NFV-enabled network slicing. The main objective of this design is to demonstrate the potential of our proposed MILP formulations and solution appraoch for the network slicing design problem. Without loss of generality, we generate the augmented virtual network based on NF requests. We take a large-scale DWDM network, DARPA CORONET CONUS [22] (illustrated in Fig. 2), as the physical network which has 75 nodes, 99 links and an average nodal degrees of 2.6.

Two network slicing instances are generated. For the single network slice construction, demands with NF requests are randomly generated, which have node ratios 50%, 60%, 70%, 80%, and 90% over the number of physical nodes, respectively. For single network slice construction, we report the total number of NF requests in Table III and call them testing cases 1–5. In Table III, we let "NodeRatio" be the virtual to physical node ratio, and "NF Node#" and "NF Req#" represent the number of virtual nodes and links with NF requests in the testing cases, respectively. For network packing problem, we consider two network slices scenario. Virtual network and NF requests of each network slice is with half node number (by ceiling) of testing cases 1-3, respectively.

Case	NodeRatio	NF Node#	NF Req#
Case 1	50%	38	61
Case 2	60%	45	152
Case 3	70%	53	172
Case 4	80%	60	194
Case 5	90%	68	222

TABLE III
Parameters in testing cases for single network slice construction

We consider three types of network functions, denoted as set $\mathcal{F} = \{1, 2, 3\}$, which are supported by all physical nodes.

Fig. 2. CONUS network [22]

	CPNode	PNode Ratio	PNode Capa	NodeCapa Ratio	NF Ratio	Flow Ratio
Case 1	30	40.0%	7415	45.2%	20.6%	24.6%
Case 2	43	57.3%	9733	59.3%	24.1%	35.2%
Case 3	53	70.7%	11516	70.2%	25.9%	44.3%
Case 4	53	70.7%	12405	64.9%	26.4%	38.5%
Case 5	56	74.7%	13227	69.2%	29.2%	40.0%

TABLE IV
Physical network resource consumption by a single network slice

Initially, physical node and link capacities are randomly generated following uniform distribution over the range $[150, 300]$. The number of NF instances requested for each demand is randomly generated following uniform distribution over the range $[1, 4]$. We assume that each instance of a NF consumes a unit of physical capacity.

For the single network slice construction, we report the resource consumption in the physical network with NF requests in Table IV, where "CPNode" denotes the utilization of physical nodes for NF instance deployment, "PNode Ratio" is the ratio of utilized physical nodes over all physical nodes, "PNode Capa" represents the total physical resource consumption in unit. "NodeCapa Ratio", "NF Ratio", and "Flow Ratio" are the ratios of "PNode Capa", resource consumption of NF deployment and routing over the node capacity of corresponding physical nodes, respectively. In testing cases 4 and 5, no feasible solutions can be produced with the generated NF requests and initial physical node capacity. We enlarge the physical node capacity following uniform distribution over the range of $[150, 350]$ to obtain feasible solutions. After acquiring the solutions for single network slice construction testing cases, we observe that (1) given sufficient physical node and link capacities, our approach can produce optimal solutions for the single network slice construction problem, (2) though the resource consumption on physical nodes is increasing (in terms of "PNode Capa"), the values are not linearly increasing with respect to either "CPNode" or "PNode Ratio", and (3) when we analyze "NodeCapa Ratio" and further divide it into "NF Ratio" and "Flow Ratio", the resource consumption of

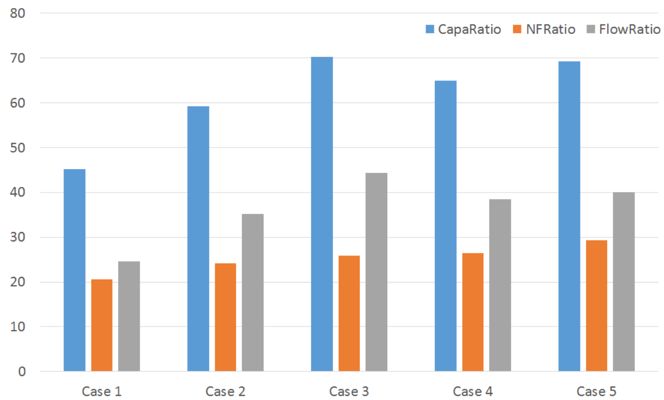

Fig. 3. Physical node resource consumption by a single network slice

	CPNode	PNode Capa	NodeCapa Ratio	NF Ratio
Case 1	30	7415	43.3%	20.4%
Case 2	43	9733	51.2%	24.3%
Case 3	53	11516	58.2%	27.7%

TABLE V
PHYSICAL NODE RESOURCE CONSUMPTION BY TWO NETWORK SLICES

NF deployment increases steadily with "NF Req#", while the resource consumption on routings/flow does not grow monotonically. The observations (2) and (3) actually demonstrate that our MILP approach can help minimize the resource consumption through generating proper paths for routings and controlling NF deployment whenever possible. We also illustrate the trends on physical node resources consumption in Fig. 3. For all our testing cases, the computational time is within 5 minutes which also shows the effectiveness of our approach.

We next report the physical node resource consumption by two network slices. Table V demonstrates that two network slices consumes less node resources on request realization and slightly more extra resources on NF deployments, maximally about 6.9%, compared with single network slice with same node number. The overall node resource consumption is saved up to 10.6% (in Case 3).

VI. CONCLUSION

In this paper, we study the NFV-enabled network slicing problem. We propose the formal definition for single network slice construction, and network slicing supporting multiple slices. We solve the network slicing problem by network packing problem, which takes network slice candidates as inputs. Corresponding mathematical formulation are presented. Leveraging formulations of network packing and single network slice construction, we present heuristic-based column generation as solution approach. We also demonstrate the extendability of proposed approach for more generalized network slicing with multiple types of VNFs.

REFERENCES

[1] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE access*, vol. 3, pp. 1206–1232, 2015.

[2] ETSI, "Network function virtualization: Infrastructure overview," ETSI, Tech. Rep. ETSI GS NFV-INF 001, 2015.

[3] L. Liu, R. Muñoz, R. Casellas, T. Tsuritani, R. Martínez, and I. Morita, "OpenSlice: An OpenFlow-based control plane for spectrum sliced elastic optical path networks," *Optics express*, vol. 21, no. 4, pp. 4194–4204, 2013.

[4] K. Nakauchi, K. Ishizu, H. Murakami, A. Nakao, and H. Harada, "Amphibia: A cognitive virtualization platform for end-to-end slicing," in *Communications (ICC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–5.

[5] "Study on new services and markets technology enablers," 3GPP, Tech. Rep. 22.891, 2.

[6] "View on 5G architecture," 5G PPP Architecture Working Group, Tech. Rep., Jul. 2016.

[7] "Network slicing for 5G network & services," 5G America, Tech. Rep., 2016.

[8] M. Gramaglia, I. Digon, V. Friderikos, D. von Hugo, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Flexible connectivity and qoe/qos management for 5g networks: The 5g norma view," in *Communications Workshops (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 373–379.

[9] C. J. Bernardos, O. Dugeon, A. Galis, D. Morris, C. Simon, and R. Szabó, "5g exchange (5gex)–multi-domain orchestration for software defined infrastructures," *focus*, vol. 4, no. 5, p. 2, 2015.

[10] N. Carapellese, M. Tornatore, and A. Pattavina, "Energy-efficient baseband unit placement in a fixed/mobile converged wdm aggregation network," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 8, pp. 1542–1551, 2014.

[11] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal bbu placement for 5g c-ran deployment over wdm aggregation networks," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1963–1970, 2016.

[12] M. Ghaznavi, A. Khan, N. Shahriar, K. Alsubhi, R. Ahmed, and R. Boutaba, "Elastic virtual network function placement," in *Cloud Networking (CloudNet), 2015 IEEE 4th International Conference on*. IEEE, 2015, pp. 255–260.

[13] M. Xia, M. Shirazipour, Y. Zhang, H. Green, and A. Takacs, "Network function placement for NFV chaining in packet/optical datacenters," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1565–1570, 2015.

[14] N. M. K. Chowdhury and R. Boutaba, "Network virtualization: state of the art and research challenges," *IEEE Communications magazine*, vol. 47, no. 7, 2009.

[15] M. Chowdhury, M. R. Rahman, and R. Boutaba, "Vineyard: Virtual network embedding algorithms with coordinated node and link mapping," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 1, pp. 206–219, 2012.

[16] X. Foukas, M. K. Marina, and K. Kontovasilis, "Orion: Ran slicing for a flexible and cost-effective multi-service mobile network architecture," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 2017, pp. 127–140.

[17] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5g and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[18] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on ran: Flexibility and resources abstraction," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, 2017.

[19] H. C. Bahl, "Column generation based heuristic algorithm for multi-item scheduling," *AIIE Transactions*, vol. 15, no. 2, pp. 136–141, 1983.

[20] D. Cattrysse, M. Salomon, R. Kuik, and L. N. Van Wassenhove, "A dual ascent and column generation heuristic for the discrete lotsizing and scheduling problem with setup times," *Management Science*, vol. 39, no. 4, pp. 477–486, 1993.

[21] M. Mourgaya and F. Vanderbeck, "Column generation based heuristic for tactical planning in multi-period vehicle routing," *European Journal of Operational Research*, vol. 183, no. 3, pp. 1028–1041, 2007.

[22] DARPA, "CORONET continental united states (CONUS) topology," http://www.monarchna.com/topology.html.