Robust Virtual Network Function Provisioning Under Random Failures on Network Function Enabled Nodes

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*—Network function virtualization enables on-demand network function (NF) deployment providing agile and dynamic network services. Early works on NF focused on its provisioning, design, and management with quality metrics – NF-service availability and reliability under system failure(s). To facilitate flexible NF service recovery and migration with high reliability against random NF-enabled node failures, with a known NF resource pool, we first introduce a new NF service evaluation metric to quantify the minimum reliability among all requested NFs for all end-to-end demands – a tight lower bound on individual NF's service reliability among all requests. We then study the robust virtual network function (VNF) provisioning problem where only a limited number of VNF instances may be instantiated while maximizing the proposed evaluation metric. We present exact solution approach which guarantees the minimum reliability of all NF service to be in the range [76%, 94%] even when physical nodes may fail with a very high (50%) probability.**

***Index Terms*—Robustness of network function service, network function virtualization, cross-layer network, Quality-of-Service (QoS).**

I. Introduction

Network function virtualization (NFV) allows network functions (NFs) to be realized as on-demand services without deploying costly proprietary networking hardware, which serves as a building block supporting the key features and development of the fifth Generation (5G) telecommunication systems [1]. Through NFV, physical resources can be allocated or reallocated to instances of VNFs, thus it not only provides more flexibility but also shortens the enabling time of new NF services [2]. NF recovery and migration are the major approaches to guarantee the continuity, resilience, and security of NF services [3–5]. When VNF manager/orchestrator [6] cannot reach NF instances, it initiates the fail-over to other available NF instances and automatically recovers NF services and/or instantiates new VNFs [7]. Meanwhile, dynamic and flexible VNF migration also reduces power consumption and the burden on hardware capacity [3].

In NFV, end-to-end demands are realized through network flows passing through a series of network functions (with or without a specific sequence) deployed onto NF-enabled containers/servers in the physical infrastructure [8], [9]. "Service function chaining" (SFC) and "non-chained" NFs are commonly used to denote the deployment of the required NFs with or without a sequential order, respectively [10]. Though SFC poses more strict requirements on resource provisioning and dependencies among NF instances, in this paper we unveil a robust and generalized VNF provisioning approach which is suitable for both non-chained NFs and SFC. We adopt cross-layer network structures to model the logical demands and the underlying network infrastructure.

Network infrastructure failures and system attacks threaten the performance, resilience, and security of NF services [11] and the underlying telecommunication networks, which motivate extensively studies on network reliability under a single-layer network setting [12–14]. The reliability analysis of a cross-layer network [15–17] imposes more challenges than its single-layer counterpart as it evaluates the probability of the logical network to remain connected after (random) physical link failure(s). Cutsets [18–20] and spanning trees [21–23] are the two typical network structures utilized to estimate network reliability.

ETSI [24], a leading investigation agency of NFV, defines the end-to-end NF reliability/availability to be the probability that NF "components have not failed after a time period" with known NF-instance locations. With this definition, if both NF deployment and network component failure probabilities are known a priori, Casazza et al. [25] showed that the best (fractional) NF assignment guaranteeing high availability can be determined through backup VNFs. In [26], a neural network-based machine learning algorithm was proposed which exploits the information of VNF forwarding graph to predict future resource requirement. With known physical link failures and backup resources, Soualah et al. [27] used Weibull distribution in formulating the meantime between failures and proposed a decision tree approach targeting the full recovery of NF services. Ding et al. [28] determined backup VNFs and their deployment considering the resources and reliability of the physical nodes. Bijwe et al. [29] assigned priorities to important VNFs to reduce NF service downtime.

While the above studies provide valuable insights from different aspects of NF services, they cannot be used to quantify system capability and reliability to support NF recovery and migration [30–32] as well as seamless NFV state transitions [33], [34] under component failure(s). In this paper we propose an evaluation metric on robust VNF provisioning which measures the *minimum NF reliability* (the tight lower bound of NF reliability) among all required NFs in the NF resource pool for all demands, which is also suitable for the SFC requests. Our goal is to provide VNF managers/orchestrators a way to evaluate the strategies to instantiate VNFs on available NF-enabled nodes (NF resource pools) based on the information of the physical infrastructure and resource utilization. We further study a robust VNF provisioning problem with the objective to maximizing the proposed evaluation metric with limited number of VNFs [28] instantiated from the NF resource pools to reduce redundancy and improve physical resources utilization.

Our contributions are summarized as follows. (1) Given NF resource pools and failure probabilities on physical nodes, we propose an evaluation metric on the reliability among all NFs, which is suitable for both non-chained NFs and SFC. (2) We present the robust NF provisioning problem which determines the minimum NF instantiation with a guaranteed low bound on the reliability of NF services. (3) We propose an exact solution approach based on mixed-integer programming. (4) Our work may be extended to the robust design and analysis of interdependent systems (integrated with facility location).

The rest of the paper is organized as follows. In Section II, we introduce the evaluation metric and define the robust VNF provisioning problem, followed by its solution approach in Section III. The experimental settings and simulation results are presented in Section IV. We conclude our work in this paper and future research directions in Section V.

II. Notations and Problem Description

In this section, we first provide the general notations used in the discussions. We then propose the robust NF-service evaluation metric and study the robust VNF provisioning problem which minimizes the number of instantiated VNFs while maximizing the robust NF-service evaluation metric. Let $G_P = (V_P, A_P)$ [or $G_P = (V_P, E_P)$] be the physical infrastructure with node set V_P and arc set A_P [or edge set E_P]. Let $G_L = (V_L, E_L)$ denote the logical network composed of end-to-end service requests $\mathcal{D} = \{d_{st}\}, s, t \in V_L, (s,t) \in E_L$, and the required NF set F. Let node set $V_P^f \in V_P$ denote a physical resource pool for NF f (candidate physical nodes to deploy f) and $V_P^F = \cup_{f \in F} V_P^f$ be the NF-enabled node set. Each of the NF-enabled nodes is with failure probability ρ_i, $i \in V_P^F$ and $0 \leq \rho_i \leq 1$.

We assume that the NF requests $d_{st}, s, t \in V_L$ are known a priori. Let d_{st} be a tuple $[(s,t), \sigma_{st}, F_{st}]$, where σ_{st} indicates whether the request is with SFC or not; if yes, $\sigma_{st} = 1$, otherwise $\sigma_{st} = 0$. We let $\widetilde{\mathcal{P}}$ and $\overrightarrow{\mathcal{P}}$ be the undirected and directed path sets in the physical network. A demand d_{st} with NF requests is *fulfilled* if it is routed through path $p_{st} \in \overrightarrow{\mathcal{P}}$ visiting all required NFs in the sequence defined in SFC when $\sigma_{st} = 1$, or otherwise routed through undirected path $\eta_{st} \in \widetilde{\mathcal{P}}$ visiting all required NFs with $\sigma_{st} = 0$. To simplify the notation, we let $\mathcal{P}$ represent the path set containing all undirected and directed paths of all NF requests.

Notations and parameters are summarized in Table I.

Notation	Description
$G_P = (V_P, A_P)$, $G_P = (V_P, E_P)$	Physical infrastructure with V_P as its node set and A_P and E_P as its arc and edge sets, respectively
$G_L = (V_L, E_L)$	Logical network with V_L, E_L as its node and edge sets, respectively
i	$i \in V_P$ denotes a physical node
$s, t, (s,t)$	$s, t \in V_L$ denote the logical nodes, and $(s,t) \in E_L$ represents the logical edges
$\mathcal{P}$, η_{st}, p_{st}	$\mathcal{P}$ represents the path sets in G_P, where $\eta, p \in \mathcal{P}$ denote the undirected and directed paths, respectively. η_{st}, p_{st} indicate paths between s and t, and $\mathcal{P}_{st} = \{\eta_{st}, p_{st}\}$
F, F_{st}, f	F represents all the NFs; F_{st} denotes the required NFs for the end-to-end demand over (s,t); and $f \in F$ is a network function
V_P^F	A set of all NF-enabled nodes, $V_P^F \subseteq V_P$
$\Gamma(F), n_i^f$	$\Gamma(F) = [\{f, i, n_i^f\} : f \in F, i \in V_P]$ is the deployment of NF instances, where n_i^f is the instances of f deployed onto i
Parameter	**Description**
$\mathcal{D}, d_{st}, \sigma_{st}$	$\mathcal{D}$ is a set of service requests, where each $d_{st} \in \mathcal{D}$ is a tuple $[(s,t), \sigma_{st}, F_{st}]$ representing the end-to-end service request between s and t; F_{st} is the set of required NFs for demand d_{st}, where $\sigma_{st} = 1$ if d_{st} is SFC (i.e., $f \in F_{st}$ should be executed in a fixed sequence), or otherwise $\sigma_{st} = 0$
ρ_i	The failure probability of NF-enabled physical node $i, i \in V_P^F$

TABLE I
NOTATIONS AND PARAMETERS

A. Robust NF Service Evaluation Metric

Our robust NF-service evaluation metric is based on the following observations.

Observation 1: Given an NF-enabled node pool V_P^F and requests $\mathcal{D} = \{d_{st}\}$, where request d_{st} is realized through a path η_{st}. d_{st} *cannot* be fulfilled if and only if $V_P^f \cap \eta_{st} = \emptyset$, $f \in F_{st}$.

Observation 1 is derived from the fact that d_{st} can only be fulfilled if and only if (all) the required NFs are deployed onto physical node(s) in its selected path η_{st}.

If $F_{st} = \{f\}$ and η_{st} is chosen for all d_{st}, $Prob(d_{st})$, the probability of d_{st} being fulfilled, is then $(1 - \prod_{i \in V_P^f \cap \eta_{st}} \rho_i)$. We now consider a more generalized setting where demands are with single or multiple NFs and their routings η_{st} are not selected (but with candidates $\mathcal{P}_{st}$).

Definition 1: Given NF-enabled node pool V_P^F, the *robust NF-service evaluation metric*, denoted as $\mathcal{RP}(d_{st})$, is

$$\mathcal{RP}(d_{st}) = \min_{f \in F_{st}} \max_{\eta_{st} \in \mathcal{P}_{st}} \left[1 - \prod_{i \in \Gamma(f) \cap \eta_{st}} \rho_i \right].$$

Note here that d_{st} with multiple non-chained NF requests is fulfilled if and only if all required NFs are satisfied. Thus, the robust evaluation metric $\mathcal{RP}(d_{st})$ is determined by the worst best-case scenario among all requested NFs realized through the best-known paths in $\mathcal{P}_{st}$. Hence, $\mathcal{RP}$ provides an *estimated low bound* of NF-service reliability for all demands.

Different from the non-chained NF requests, SFC requests are fulfilled only when all required NFs are served in a specified sequence. Without loss of generality, we assume that (1) *the same NF request will not be fulfilled more than once on different NF-enabled nodes*, and (2) *each NF-enabled node will not carry out multiple NF requests in SFC*.

Definition 2: Given NF-enabled node pool V_P^F, the robust NF evaluation metric of SFC request d_{st} is

$$\mathcal{RP}(d_{st}) = \min_{f \in F_{st}} \max_{p_{st} \in \mathcal{P}_{st}} \left[1 - \prod_{i \in \Gamma(f) \cap p_{st}} \rho_i \right] / |F_{st}|!.$$

Since demands with SFC request are fulfilled only when all requested NFs are deployed onto p_{st} and visited in a predefined sequence, there is only one valid case out of $|F_{st}|!$ permutations. $\mathcal{RP}(d_{st})$ is then determined by the worst best-case scenario among all requested NFs realized through the best-known paths in $\mathcal{P}_{st}$ (with the highest probability to survive).

Considering multiple NF requests in a given NFVI (NFV infrastructure) and managed by the same NFV MANO (NFV management and organization), we define the robust NF-service evaluation metric among all NF requests as follows.

Definition 3: Given G_P, G_L, a set of NFs F, NF-enabled node pool V_P^F and node failure probability $\rho_i, i \in V_P^F$, $\mathcal{RP}(V_P^F) = \min_{d_{st} \in D} \mathcal{RP}(d_{st})$.

B. Illustrations: NF Service Reliability vs. Robust NF Service Evaluation Metric

We evaluate the robust NF-service evaluation metric via an instance illustrated in Fig. 1 and present its difference to the NF-service reliability defined in [24]. In this example, two demands with NF requests d_{12} and d_{34} are considered. Demand d_{12} requires SFC $f_1 \rightarrow f_2$ and d_{34} requires non-chained NFs $\{f_1, f_2\}$. NF-enabled nodes, their supported NFs, and their failure probabilities are labeled in Fig. 1(b). Candidate physical nodes to enable/deploy f_1's are in set $V_P^1 = \{1, 3, 4, 5\}$, and those for f_2's are in $V_P^2 = \{2, 3, 5, 6\}$. d_{12} is routed through a directed path $\{(1, 5), (5, 2)\}$, and d_{34} is routed through an undirected path $\{(4, 6), (6, 3)\}$. Based on the assumptions given in the previous section, the robust NF-service evaluation metric $\mathcal{RP}(\{d_{12}, d_{34}\}) = \min\{1-0.1, 1-0.2, (1-0.2 \times 0.1)/2, (1-0.1 \times 0.2)/2\} = 0.49$.

Different from $\mathcal{RP}(\{d_{12}, d_{34}\})$, NF-service reliability of d_{12} is

$$\begin{aligned}
&1 - Prob(f_1, f_2 \text{ both failed}) - Prob(\text{only } f_2 \text{ failed}) \\
&- Prob(\text{only } f_1 \text{ failed}) - Prob(f_1, f_2 \text{ fulfilled not in-order}) \\
&= 1 - 0.1 * 0.2 * 0.1 - 0.9 * 0.2 * 0.1 - 0.9 * 0.2 * 0.1 - 0 \\
&= 0.962.
\end{aligned}$$

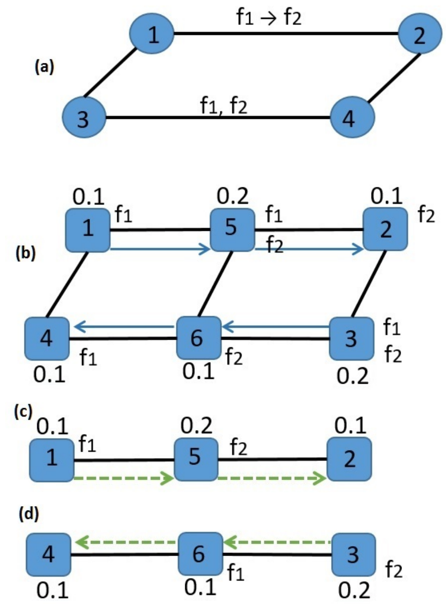

Fig. 1. NF reliability

The NF-service reliability of d_{34} is

$$\begin{aligned}
&1 - Prob(f_1, f_2 \text{ both failed}) - Prob(\text{only } f_2 \text{ failed}) \\
&- Prob(\text{only } f_1 \text{ failed})] \\
&= 1 - 0.2 * 0.1 * 0.1 - 0.2 * 0.1 * 0.9 - 0.2 * 0.9 * 0.1 \\
&= 0.962
\end{aligned}$$

The examples above show that NF-service reliability is measured when the deployment of NF instances and routings are determined. In contrast, since the robust NF-service evaluation metric already evaluates the minimum NF reliability, the routings selected and the deployment of non-chained NFs or SFC would always be better than or at least equal to the metric. In other words, the robust NF-service evaluation metric provides a tight lower bound for each NF's reliability.

This instance also shows that with the limitation imposed on the NF-enabled nodes, the selection of NF-enabled nodes also impact the robust NF-service evaluation metric. Hence, in the following section, we study the robust NF provisioning problem which aims at maximizing our proposed NF-service evaluation metric via NF-enabled node selection for NF request realization.

C. Robust NF Provisioning

When taking the failures of NF-enabled nodes into consideration, we now define the robust VNF provisioning problem. Given G_P, G_L, $\mathcal{D}$, and V_P^F. $d_{st} \in \mathcal{D}$, $s, t \in V_L$, is mapped onto a directed path $p_{st} \in \mathcal{P}$ for SFC request, or undirected path $\eta_{st} \in \mathcal{P}$ for NF request. We would like to determine a limited number of NF-enabled nodes to support each required NF and guarantee that demands are routed through their required NFs. This problem considers both non-chained NF and SFC requests.

Definition 4: Given N_f as the limited number of NF-enabled nodes supporting NF f, the **robust VNF provisioning problem** is to determine the NF deployment which maximizes the robust NF-service evaluation metric: $\max_{V_P^f:|V_P^f \leq N_f|} \mathcal{RP}(V_P^F)$.

D. Maximizing $\mathcal{RP}(V_P^F)$ via Minimizing $\mathcal{FP}(V_P^F)$

In this section, we discuss the failure probability of NF services, denoted as $\mathcal{FP}(V_P^F)$, which is the counterpart of the robust NF-service evaluation metric. We show that the robust VNF provisioning can be achieved via finding the *minimum robust NF failure evaluation metric*.

Proposition 1: $1 - \mathcal{RP}(V_P^f) = \mathcal{FP}(V_P^f)$, with $d_{st} \in \mathcal{D}$ and $f \in F_{st}$.

Proposition 1 derives directly from Definitions 1 and 2, which also holds for SFC requests. Hence, we have the following conclusion.

Theorem 1: $\max_{V_P^F} \mathcal{RP}(V_P^F) = \min_{V_P^F} \mathcal{FP}(V_P^F)$.

Based on Proposition 1 and Theorem 1, we have the following equations.

$$\mathcal{FP}(d_{st}) = \max_{f \in F_{st}} \min_{\eta_{st} \in \mathcal{P}_{st}} \left[\prod_{i \in \Gamma(f) \cap \eta_{st}} \rho_i \right] \quad (1)$$

$$\mathcal{FP}(d_{st}) = \max_{f \in F_{st}} \min_{\eta_{st} \in \mathcal{P}_{st}} \left[\prod_{i \in \Gamma(f) \cap p_{st}} \rho_i \right] / |F_{st}|! \quad (2)$$

$$\mathcal{FP}(V_P^F) = \max_{d_{st} \in D} \mathcal{FP}(d_{st}) \quad (3)$$

In the next section, we demonstrate that solving the robust VNF provisioning via minimizing NF failure evaluation metric provides a linearization of non-linear equations. We then propose the solution approach accordingly.

III. Solution Approach

In this section, we demonstrate how to utilize $\mathcal{FP}(V_P^F)$ to formulate robust VNF provisioning problem and propose the mixed-integer programming solution approach for the problem. The variables and parameters used in this section are presented in Table II.

A. Formulations for NF Request

We now present the mathematical formulations for the maximal reliable NF deployment problem based on the NF service failure probability. We first turn the non-linear objective $\min_{V_P^F} \max_{d_{st} \in \mathcal{D}} \min_{\eta_{st} \in \mathcal{P}_{st}} \Pi_{i \in V_P^f \cap \eta_{st}} \rho_i$ into its linearized counterpart

$$\min_{V_P^F} \max_{\substack{f \in F_{st} \\ d_{st} \in \mathcal{D}}} \min_{\eta_{st} \in \mathcal{P}_{st}} \sum_{i \in V_P^f \cap \eta_{st}} \ln(\rho_i) \quad (4)$$

by applying the $\ln(\cdot)$ function.

Parameter	Description
N_f	The number limitation for NF deployed locations with $f \in F$
ρ_i	The failure probability of physical node i with $i \in V_P$
$\delta^i_{\eta_{st}}$	A binary indicator showing whether physical node i is on path η_{st} or not, $\eta_{st} \in \mathcal{P}_{st}, (s,t) \in E_L$; if yes, $\delta^i_{\eta_{st}} = 1$, otherwise $\delta^i_{\eta_{st}} = 0$
γ^f_{st}	A binary indicator showing whether f is requested by d_{st} or not; if yes, $\gamma^f_{st} = 1$, otherwise, $\gamma^f_{st} = 0$
M	A very large number
Variable	**Description**
λ	The upper bound of NF failure probability of NF requests in $\mathcal{D}$
ξ^f_{st}	NF failure probability of NF $f \in F$ and $d_{st} \in \mathcal{D}$
$x_{p_{st}}$	A binary variable indicating whether path $p_{st} \in \mathcal{P}_{st}$ is selected to fulfill $d_{st} \in \mathcal{D}$
y^{if}_{st}	A binary variable indicating whether physical node i provides NF requests f for d_{st} or not; if yes, $y^{if}_{st} =$ 1, otherwise, $y^{if}_{st} = 0$
h_i	A binary variable which indicates whether a network function is deployed onto physical node i or not; if yes, $h_i = 1$, otherwise, $h_i = 0$
z^f_i	A binary variable which indicates if network function f is deployed onto physical node i; if yes, $z^f_i = 1$, otherwise, $z^f_i = 0$
β_{st}	A binary auxiliary variable which indicates if demand d_{st} is selected under the SFC setting; if yes, $\beta_{st} = 1$, otherwise $\beta_{st} = 0$

TABLE II
PARAMETERS AND VARIABLES

With Theorem 1, the formulation presented below is the robust NF evaluation metric value of NF request with (4) as the objective:

$$\min_{\lambda,x,y,z,\xi,h} \lambda$$

$$s.t. \sum_{i \in V_P} h_i \leq N_f, f \in F \quad (5)$$

$$\lambda \geq \xi^f_{st}, \quad f \in F, d_{st} \in \mathcal{D} \quad (6)$$

$$\xi^f_{st} = \sum_{i \in V_P} ln(\rho_i) y^{if}_{st}, \quad f \in F, d_{st} \in \mathcal{D} \quad (7)$$

$$y^{if}_{st} \geq z^f_i + \delta^i_{\eta_{st}} x_{\eta_{st}} + \gamma^f_{st} - 2, \quad f \in F, d_{st} \in \mathcal{D}, \eta_{st} \in \mathcal{P}_{st}, i \in V_P \quad (8)$$

$$y^{if}_{st} \leq z^f_i, \quad f \in F, i \in V_P \quad (9)$$

$$y^{if}_{st} \leq \delta^i_{\eta_{st}} x_{\eta_{st}}, \ f \in F, i \in V_P, d_{st} \in \mathcal{D}, \eta_{st} \in \mathcal{P}_{st} \quad (10)$$

$$y^{if}_{st} \leq \gamma^f_{st}, \quad f \in F, i \in V_P, d_{st} \in \mathcal{D} \quad (11)$$

$$h_i \geq z^f_i, \quad f \in F, i \in V_P \quad (12)$$

$$\sum_{\eta_{st} \in \mathcal{P}_{st}} x_{\eta_{st}} = 1, \quad d_{st} \in \mathcal{D} \quad (13)$$

$$\lambda, \xi^f_{st} \geq 0, z^f_i, y^{if}_{st}, h_i, x_{\eta_{st}} \in \{0,1\}, \eta_{st} \in \mathcal{P}_{st}, (s,t) \in E_L, f \in F, d_{st} \in \mathcal{D}, i \in V_P \quad (14)$$

Constraint (5) enforces the upper bound for the number of nodes deployed with NFs. Constraint (6) records the value of NF failure evaluation metric (linearized) among all demands for all NFs. Constraint (7) captures the robust NF failure evaluation metric value (linearized, i.e., $\ln(\rho_i)$ as in constraint (4)) of demand $d_{st} \in \mathcal{D}$ and $f \in F$. Based on Definition 1, constraint (8) determines whether f is deployed onto physical node i for demand $d_{st} \in \mathcal{D}$, where (i) $z_i^f = 1$ when f is deployed onto physical node i; (ii) $\delta^i_{\eta_{st}} = 1$ when node i deployed with an NF is on a selected path η_{st} for d_{st}; and (iii) $\gamma^f_{st} = 1$ when d_{st} requires NF f. Constraints (9) – (11) force variable y^{if}_{st} to be 0 when any of the (i) to (iii) above is not satisfied. Constraint (12) indicates whether physical node i is deployed with any NFs. Constraint (13) selects a single physical route for demand $d_{st} \in \mathcal{D}$. Constraint (14) provides feasible regions for all variables.

Note here that the variable λ in constraint (6) records the value of the robust NF failure evaluation metric achieved by NF request through ξ^f_{st}. As the objective of the reformulation is to find the minimum λ, it also encourages evaluation metric value ξ^f_{st} to be minimized. Therefore, the above reformulation solves the maximal reliable NF deployment problem.

We next present the formulation for SFC service reliability.

B. Formulations for SFC Service Reliability

Different from the non-chained NF failure probability, the **SFC failure probability** is $1 - \max_{\Gamma(F)} \min_{\substack{f\in F_{st}\\ d_{st}\in\mathcal{D}}} \max_{\eta_{st}\in\mathcal{P}_{st}} \left[1 - \Pi_{i\in\Gamma(f)\cap p_{st}}\rho_i\right] / |F_{st}|!$ with $d_{st} \in \mathcal{D}$.

Proposition 2: For requests with SFC, we have $\max_{\Gamma(F)} \min_{\substack{f\in F_{st}\\ d_{st}\in\mathcal{D}}} \max_{p_{st}\in\mathcal{P}_{st}} \left[1 - \Pi_{i\in\Gamma(f)\cap p_{st}}\rho_i\right] / |F^*_{st}|!$ $= 1 - \min_{\Gamma(F)} \max_{\substack{f\in F_{st}\\ d_{st}\in\mathcal{D}}} \min_{p_{st}\in\mathcal{P}_{st}} \Pi_{i\in\Gamma(f)\cap p_{st}}\rho_i / |F^*_{st}|!$, where F^*_{st} represents the requested NFs of $d^*_{st} = \arg\min_{d_{st}\in\mathcal{D}, f\in F_{st}} \left[\Pi_{i\in\Gamma(f)\cap p_{st}}\rho_i\right]$.

We introduce here an auxiliary variable ω_{st} which indicates whether $d_{st} \in \mathcal{D}$ is selected as the d^*_{st}. By replacing routings from undirected to directed path set (i.e., $\eta_{st} \to p_{st}$) in constraints (8), (10), (13), (14), we present the formulation for the robust SFC provisioning as follows.

$$
\begin{aligned}
&\min_{\lambda,\xi,\omega,\beta,y,x,z} \lambda \\
&\quad s.t.\ \lambda \geq \omega_{st}, && d_{st} \in \mathcal{D} && (15)\\
&\qquad \omega_{st} \geq \xi^f_{st} - \ln|F_{st}|!, && f \in F, d_{st} \in \mathcal{D} && (16)\\
&\qquad \sum_{d_{st}\in\mathcal{D}} \beta_{st} = 1 && && (17)\\
&\qquad \lambda \leq \omega_{st} + M(1-\beta_{st}), && d_{st} \in \mathcal{D} && (18)\\
&\qquad \lambda \geq \omega_{st} + M(\beta_{st}-1), && d_{st} \in \mathcal{D} && (19)\\
&\qquad \omega_{st} \geq 0, \beta_{st} \in \{0,1\}, d_{st} \in \mathcal{D} && && (20)\\
&\qquad \text{Constraints (5) and (7)–(14)}
\end{aligned}
$$

Constraint (15) is to guarantee the lower bound based on the $\mathcal{FP}$ (linearized). The newly introduced constraint (16) is used to capture the corresponding SFC request $d_{st} \in \mathcal{D}$. Constraint (17) guarantees that exactly one demand $d_{st} \in \mathcal{D}$ should be selected as the d^*_{st} which provides the $\mathcal{FP}(d_{st})$. Constraints (18) and (19) guarantee $\lambda = \omega_{st}$ for the selected d^*_{st} (when $\beta_{st} = 1$).

IV. Simulations

We present in this section the experimental design, testing case construction, and simulation results to validate the proposed solution approaches. Our goals are to demonstrate that (1) the NF-service reliability increases monotonically when the failure probability of physical nodes decreases. Especially, when the physical nodes are with failure probability less then 5%, the NF deployment and the logical-to-physical paths generated/selected by the proposed approaches can guarantee high NF reliability (greater than 90%). (2) To guarantee a high NF service reliability (say 90%), the average number of physical nodes to be deployed with NF f is small even when the physical nodes may fail with high probability in the range of [30%, 50%].

A. Setup

We select NSF network as the physical network illustrated in Fig. 2, which has 14 nodes and 21 links. NF requests are based on node pairs whose mappings onto physical nodes are known a priori. Six pairs of NF requests are constructed and listed as follows: (1,2), (1,4), (2,3), (3,5), (4,7), and (6,7). NF requests for logical arcs/links are randomly assigned with up to three NFs.

We consider that physical nodes are with random failure probabilities, where the means of these probabilities are in the range of 1% to 49% and the variance is 0.001. For each of the failure probabilities, we generate 25 testing samples and report their average as the results. For the simulations of the maximal reliable NF deployment problem, we first create testing cases which restrict the number of NF-enable nodes to be 40%, 50%, and 60% of the physical nodes.

Based on the settings above, two sets of testing cases are created. The first testing cases for the maximal NF reliable deployment problem have (i) NSF as the physical network, (ii) demands with up to three randomly assigned NF requests, (iii) a given limitation on the number of NF deployed locations, and (iv) random node failure probability. The proposed setting is to verify that when the number of NF locations decreases,

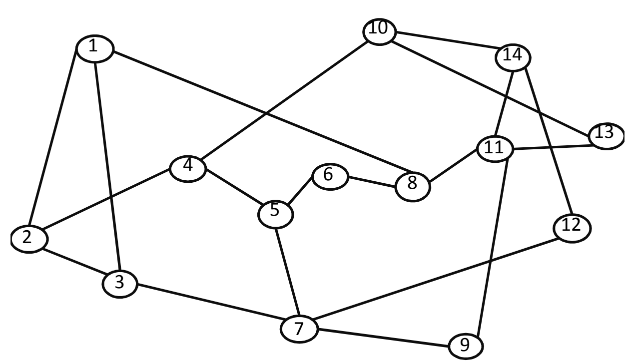

Fig. 2. NSF

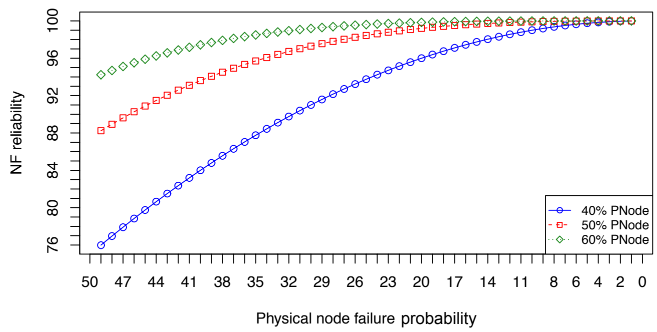

Fig. 3. NF service reliability

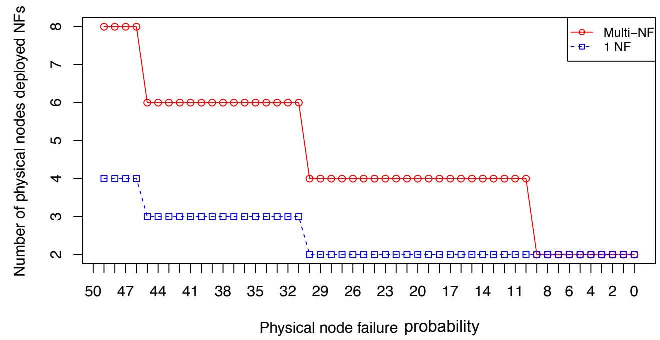

Fig. 4. NF service reliability vs. NF deployment

whether the NF service reliability also goes down corresponding. Meanwhile, when the node failure probability increases, whether the NF service reliability also decreases.

The second testing cases have (i) a fixed NF service reliability (90%), and (ii) random physical node failure probability. The purpose of the setting is again to evaluate that with a fixed NF service reliability, when the node failure probability increases, whether extra NF-deployed nodes are required to fulfill the requirement of the service level.

We report the simulation results with the two sets of testing cases in the following section.

B. Simulation Results

The simulation results for the maximal NF reliable deployment problem are presented in Fig. 3. The three lines in blue, red, and green colors represent the testing cases with 40%, 50%, and 60% of NF-enabled physical nodes. The x-axis represents the physical node failure probability (in mean value) and the y-axis denotes the NF service reliability (in percentage). Each plotted node/dot in the figure presents the average NF service reliability for all testing samples. With up to 50% failure probability of the NF-enabled nodes, the NF reliability reaches 75%. When the number of NF-enabled nodes increases, the NF reliability increases to 87.5%, and 93.7%, respectively. We confirm our analysis that with the limitation on the number of NF-enabled nodes, the NF service reliability increases when physical node failure probability decreases. Also, given the same physical node failure probabilities, we observe that when the number of NF-enabled nodes (in terms of the mean value) decreases, the reliability of the NF service decreases as well.

Figure 4 illustrates the number of NFs deployed to reach the required level of the NF service reliability (based on the maximal number of NF-enabled nodes in the testing cases) with single NF and multiple NFs (in our testing cases, three required NFs) in each demand. To reach the fixed (90%) NF service reliability, the number of physical nodes deployed with NFs is only doubled when the number of required NFs for each demand goes from one to three even with high failure probability (10 – 50%) on physical nodes. The figure demonstrates a clear pattern between the number of nodes deployed with NFs and the NF service reliability.

In the simulation results, we observe that the NF service reliability is higher with more physical nodes deployed with the required NFs, and obviously, a lower average node failure probability leads to a higher NF service reliability under the failure(s) of physical nodes. The observations on these simulations are as expected and demonstrate the relationship between the number of NF-deployed nodes (cost-related restriction) and NF service reliability (service level).

V. Conclusion and Future Works

We studied the reliable NF deployment problems under random physical node failure in this paper. We proposed an evaluation metric, the NF service reliability, to quantify and indicate the probability of the required NFs being fulfilled for the end-to-end demands. Utilizing this evaluation metric, we studied the maximal NF reliable deployment problem and proposed the exact solution approaches to solve the problem. We designed and conducted simulations to confirm our analysis on the reliable NF deployment approaches.

In the further research, we would also like to consider physical node capacity and NF deployment costs, and evaluate the costs to introduce more NF-enabled nodes in the physical network. We also like to investigate the scenarios of shared risk group failure(s) and physical link failure(s) and their impacts on NF service reliability. Last, but not least, another research direction is to relax the assumptions on independent node failures, the correlations among NF-enabled node failures, and study their impacts on NF service reliability.

References

[1] ETSI, "Network functions virtualisation: White paper on NFV priorities for 5G," ETSI, Tech. Rep., 2017.

[2] ——, "Network functions virtualisation (NFV): Infrastructure overview," ETSI, Tech. Rep., 2015.

[3] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236–262, 2016.

[4] V. Eramo, E. Miucci, M. Ammar, and F. G. Lavacca, "An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures," *IEEE/ACM Transactions on Networking*, vol. 25, no. 4, pp. 2008–2025, 2017.

[5] W. Wang, Y. Liu, Y. Li, H. Song, Y. Wang, and J. Yuan, "Consistent state updates for virtualized network function migration," *IEEE Transactions on Services Computing*, 2017.

[6] ETSI, "Network functions virtualisation (NFV): Management and orchestration," ETSI, Tech. Rep. ETSI GS NFV-MAN 001, 2014.

[7] L. Xia, Q. Wu, D. King, and H. Yokota, "Vnf pool use cases," IETF, Tech. Rep. draft-king-vnfpool-mobile-use-case-02, 2014.

[8] B. Addis, D. Belabed, M. Bouet, and S. Secci, "Virtual network functions placement and routing optimization," in *Cloud Networking (CloudNet), 2015 IEEE 4th International Conference on*. IEEE, 2015, pp. 171–177.

[9] T. Lin, Z. Zhou, M. Tornatore, and B. Mukherjee, "Demand-aware network function placement," *Journal of Lightwave Technology*, vol. 34, no. 11, pp. 2590–2600, 2016.

[10] 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.

[11] W. Yang and C. Fung, "A survey on security in network functions virtualization," in *NetSoft Conference and Workshops (NetSoft), 2016 IEEE*. IEEE, 2016, pp. 15–19.

[12] A. M. Shooman and A. Kershenbaum, "Exact graph-reduction algorithms for network reliability analysis," in *Global Telecommunications Conference, 1991. GLOBECOM'91*. IEEE, 1991, pp. 1412–1420.

[13] M. O. Ball, C. J. Colbourn, and J. S. Provan, "Network reliability," *Handbooks in operations research and management science*, vol. 7, pp. 673–762, 1995.

[14] M. K. Chari, T. A. Feo, and J. S. Provan, "The Delta-Wye approximation procedure for two-terminal reliability," *Operations Research*, vol. 44, no. 5, pp. 745–757, 1996.

[15] K. Lee, H.-W. Lee, and E. Modiano, "Reliability in layered networks with random link failures," *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 6, pp. 1835–1848, 2011.

[16] H.-W. Lee, K. Lee, and E. Modiano, "Maximizing reliability in WDM networks through lightpath routing," *IEEE/ACM Transactions on Networking (TON)*, vol. 22, no. 4, pp. 1052–1066, 2014.

[17] Z. Zhou, T. Lin, K. Thulasiraman, G. Xue, and S. Sahni, "Cross-layer network survivability under multiple cross-layer metrics," *IEEE/OSA Journal of Optical Communications and Networking (JOCN)*, vol. 7, no. 6, pp. 540 – 553, 2015.

[18] M. O. Ball and J. S. Provan, "Bounds on the reliability polynomial for shellable independence systems," *SIAM Journal on Algebraic Discrete Methods*, vol. 3, no. 2, pp. 166–181, 1982.

[19] R. Taylor and R. Addie, "An algorithm for calculating the availability and mean time to restore for communication through a network," *Computer Networks and ISDN Systems*, vol. 20, no. 1-5, pp. 109–114, 1990.

[20] L. D. Nel and C. J. Colbourn, "Combining Monte Carlo estimates and bounds for network reliability," *Networks*, vol. 20, no. 3, pp. 277–298, 1990.

[21] Z. Zhou, T. Lin, K. Thulasiraman, G. Xue, and S. Sahni, "Novel survivable logical topology routing in IP-over-WDM networks by logical protecting spanning tree set," in *Proc. of 4th 4th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2012, pp. 650–656.

[22] J. Yallouz, O. Rottenstreich, and A. Orda, "Tunable survivable spanning trees," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 1, pp. 315–327, 2014.

[23] Z. Zhou, T. Lin, K. Thulasiraman, and G. Xue, "Novel survivable logical topology routing by logical protecting spanning trees in IP-over-WDM networks," *IEEE/ACM Transactions on Networking (ToN)*, vol. 25, no. 3, pp. 1673 – 1685, 2017.

[24] ETSI, "Network functions virtualisation (NFV); Reliability; Report on models and features for end-to-end reliability," ETSI, Tech. Rep. V1.1.1, 2016. [Online]. Available: http://www.etsi.org/deliver/etsi_gs/NFV-REL/001_099/003/01.01.01_60/gs_nfv-rel003v010101p.pdf

[25] M. Casazza, P. Fouilhoux, M. Bouet, and S. Secci, "Securing virtual network function placement with high availability guarantees," *arXiv preprint arXiv:1701.07993*, 2017.

[26] R. Mijumbi, S. Hasija, S. Davy, A. Davy, B. Jennings, and R. Boutaba, "Topology-aware prediction of virtual network function resource requirements," *IEEE Transactions on Network and Service Management*, vol. 14, no. 1, pp. 106–120, 2017.

[27] O. Soualah, M. Mechtri, C. Ghribi, and D. Zeghlache, "A link failure recovery algorithm for virtual network function chaining," in *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on*. IEEE, 2017, pp. 213–221.

[28] W. Ding, H. Yu, and S. Luo, "Enhancing the reliability of services in nfv with the cost-efficient redundancy scheme," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[29] S. Bijwe, F. Machida, S. Ishida, and S. Koizumi, "End-to-end reliability assurance of service chain embedding for network function virtualization," in *Network Function Virtualization and Software Defined Networks (NFV-SDN), 2017 IEEE Conference on*. IEEE, 2017, pp. 1–4.

[30] H. Hawilo, M. Jammal, and A. Shami, "Orchestrating network function virtualization platform: Migration or re-instantiation?" in *Cloud Networking (CloudNet), 2017 IEEE 6th International Conference on*. IEEE, 2017, pp. 1–6.

[31] V. Eramo, M. Ammar, and F. G. Lavacca, "Migration energy aware reconfigurations of virtual network function instances in nfv architectures," *IEEE Access*, vol. 5, pp. 4927–4938, 2017.

[32] B. Han, V. Gopalakrishnan, G. Kathirvel, and A. Shaikh, "On the resiliency of virtual network functions," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 152–157, 2017.

[33] L. Nobach, I. Rimac, V. Hilt, and D. Hausheer, "Statelet-based efficient and seamless nfv state transfer," *IEEE Transactions on Network and Service Management*, vol. 14, no. 4, pp. 964–977, 2017.

[34] Y. Wang, G. Xie, Z. Li, P. He, and K. Salamatian, "Transparent flow migration for nfv," in *Network Protocols (ICNP), 2016 IEEE 24th International Conference on*. IEEE, 2016, pp. 1–10.