Towards Robust Network Slice Design Under Correlated Demand Uncertainties

Andreas Baumgartner*, Thomas Bauschert*, Fabio D'Andreagiovanni†§, Varun S. Reddy*
†French National Center for Scientific Research (CNRS), France
§Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc UMR 7253, CS 60319, 60203 Compiègne, France
Email: d.andreagiovanni@hds.utc.fr
*Technische Universität Chemnitz, Chair of Communication Networks, 09126, Chemnitz, Germany
Email: andreas.baumgartner | thomas.bauschert | srva@etit.tu-chemnitz.de

***Abstract*—Network Slicing is envisaged as a key component to address the challenges arising in next generation networks concerning the deployment, control and management of services. Besides, it promotes concurrent operation of multiple logical networks with diverging requirements on a common substrate platform. In this regard, the problem of designing individual logical network slices and mapping them onto the underlying substrate network gains significance. We denote this problem as the Network Slice Design Problem. In this work, we first consider the general network slice design problem. Adopting the robust optimisation approach of Bertsimas and Sim [1], [2], we then develop two additional formulations: i) to handle traffic demand uncertainties, and ii) to account for the correlations among the uncertain traffic demands. Finally, we present an extensive evaluation of the proposed formulations using realistic network instances.**

I. Introduction

Driven by the need to enable a connected society, the Internet is currently witnessing rapid proliferation. Cisco's Visual Networking Index forecasts a threefold increase in the global IP traffic over the coming 5 years [3]. Such a growth in traffic volume is meant to exert a lot of stress on the performance of the underlying network infrastructure, mainly composed of legacy hardware-based network appliances. In order to meet the increasing QoS requirements of emerging traffic-intensive network services, telecom service providers (TSPs) have to spend large investments on their network infrastructures. Network Softwarisation is anticipated to act as a catalyst to transform the present-day telecommunication networks into cloud-based networks where virtual software-based network functions replace the legacy hardware-based network elements. It further enables the TSPs to build and host a multitude of diverse logical networks on a common substrate network infrastructure thereby promoting new business opportunities in the ICT sector.

Considering this, the NGMN alliance puts forth the concept of *Network Slicing* which conforms with the telecom service providers' view of multi-tenancy in next generation networks [4]. Network Slicing provides a basis for optimally partitioning the substrate network resources among different virtual network slices with the purpose of enabling independent control over the resources allocated to the respective slices. Each logical slice corresponds to an abstraction of a subset of physical substrate network resources tailored to meet customer-specific QoS requirements. While the topic of network slicing is well investigated in the context of wireless networks (see [5] for an overview), its extension to fixed networks is currently limited to partitioning the substrate network resources to match the resource requirements of network slice requests with pre-defined topologies as in the Virtual Network Embedding (VNE) problem addressed in [6], [7], [8]. We remark that the problem statement for network slicing must not be merely confined to the assignment of resources to pre-defined network slice requests but must also encompass the design of the individual network slices as well (i.e., the network slice topology, the number of required virtual functions, their dimensioning, and the interconnections). We refer to this problem as the *Network Slice Design Problem* (NSDP).

Public safety and emergency services that form an integral part of our day-to-day lives are expected to withstand the impacts of unpredictable events and provide robust communication services during times of distress. Usually, the traffic demand in the aforementioned circumstances is much higher compared to normal situations due to the large number of users simultaneously accessing the services. Thus, network slices supporting lifeline communications for emergency prediction, disaster relief etc., must be robustified in order to handle the unprecedented rise in traffic demand following unforeseen natural disasters [4]. To provide robust communication services in such situations, the design of such network slices has to consider the information about traffic demand fluctuations/parametric deviations. Robust Optimisation has emerged as an important mathematical tool for network planners to take into account the uncertain nature of traffic demands [9], [10], [11]. While most of the work concerning robust network planning assumes independence among the end-to-end traffic demand variations, for scenarios such as lifeline communications, it is very likely that they are spatially correlated. Conse-

This work was supported by Celtic-Plus cluster project SENDATE Secure-DCI (ID C2015/3-1).

978-1-5386-3180-5/18/$31.00 ©2018 IEEE

quently, the knowledge about correlation among the uncertain traffic demands should be included in the problem formulation. Ignoring this aspect may compromise the operation of critical public safety services. Some of the contributions that address the aspect of correlation among parametric uncertainties include robust portfolio selections [12], demand response models for energy consumption scheduling [13], robust energy and reserve scheduling [14], among others.

In this paper, we first consider the general network slice design problem proposed in [15]. Following the robust optimisation approach of Bertsimas and Sim [1], [2], we develop two different robust formulations: i) an optimisation model to handle uncertainties in traffic demands, and ii) an alternate optimisation model that builds on the previous model by considering the inclusion of spatial correlation among the uncertain traffic demands in the proposed formulation . For each of these problem formulations, we first present a simple exponential model which can be solved using separation routines. We then derive a compact reformulation of the exponential model by exploiting the property of LP duality.

The paper is structured as follows: In Section II, we formally introduce the general network slice design problem. In Section III, we develop robust network slice design models to handle traffic demand uncertainties. In Section IV, we present an extensive computational evaluation of the proposed approaches using realistic network topologies from SNDlib [16]. Section V concludes the paper summarising the findings and outlining possible directions for future work.

II. The Network Slice Design Problem

A network slice can be fundamentally described as a self-contained collection of traffic flows traversing an ordered set of service-specific virtual functions within a physical substrate network infrastructure, intended to act as an autonomous network. In the view of this, we formalise the general capacitated network slice design problem. The physical substrate network is represented by an undirected graph $G(N_s, L_s)$ composed of physical substrate nodes, and substrate links interconnecting them, with $\bar{\kappa}_{n_s} \geq 0$ for all $n_s \in N_s$, and $\bar{\kappa}_{l_s} \geq 0$ for all $(n_{s_1}, n_{s_2}) = l_s \in L_s$ denoting their respective residual capacities. The costs per unit occupancy of the substrate node and link resources are indicated by $\bar{c}_{n_s} \geq 0$ resp. $\bar{c}_{l_s} \geq 0$. In the event of insufficient spare resources in the substrate network, the node and link resources can be expanded in discrete steps of size $\kappa_{n_s} \in \mathbb{Z}_+$ resp. $\kappa_{l_s} \in \mathbb{Z}_+$, incurring a fixed cost of $c_{n_s} \geq 0$ resp. $c_{l_s} \geq 0$.

For a network slice to establish an end-to-end service for a set T of point-to-point demands, the traffic volume $d_t \in \mathbb{R}_+$ for all $t \in T$ must traverse from source $i(t)$ to destination $e(t)$, an ordered set of service-specific virtual functions $N_v(t)$. The order of precedence - i.e., the logical interconnections between the virtual functions is indicated by $(n_{v_1}, n_{v_2}) \in L_v(t)$ and can be derived from the corresponding service graph [17]. These virtual functions can be instantiated on substrate nodes and consume substrate resources when processing/routing the traffic demands. They must thus be dimensioned according to the volume of traffic routed through them. However, restrictions regarding their placement have to be taken into consideration which may arise due to technological, economic, geographical or security constraints. Accordingly, we introduce a parameter $\varphi_{n_s}^{n_v} \in \{0,1\}$ indicating the ability of the substrate node n_s to host the respective virtual function n_v. Similarly, virtual IP links must be established (between substrate nodes hosting the virtual functions) and consume substrate link resources proportional to the volume of the traversing traffic.

Decision variables $x_{n_s}^{t,n_v} \in \mathbb{R}_{[0,1]}$ indicate the fraction of demand $t \in T$ processed/routed through a virtual function $n_v \in N_v(t)$ residing on a substrate node $n_s \in N_s$. Flow variables $f_{l_s}^{t,l_v} \in \mathbb{R}_{[0,1]}$ indicate the fraction of demand $t \in T$ between the virtual functions $(n_{v_1}, n_{v_2}) \in L_v(t)$ routed over the physical link $l_s \in L_s$. Variables $y_{n_s}^{n_v} \in \mathbb{Z}_{\geq 0}$ specify the number of capacity modules of size $\kappa_{n_v} \in \mathbb{Z}_+$ assigned to the virtual function n_v residing on the substrate node n_s. Variables $u_{n_s} \in \mathbb{R}_{\geq 0}$ and $u_{l_s} \in \mathbb{R}_{\geq 0}$ indicate the utilisation of the substrate node and link resources, respectively. Let $y_{n_s} \in \mathbb{Z}_{\geq 0}$ and $y_{l_s} \in \mathbb{Z}_{\geq 0}$ denote the number of capacity modules of size κ_{n_s} and κ_{l_s} that have to be additionally installed at the substrate nodes and links, respectively to accommodate the network slice in case of insufficient substrate network capacity.

In the following, we outline the formulation of the general network slice design problem as a mixed integer program:

$$\min \sum_{n_s \in N_s} \bar{c}_{n_s} u_{n_s} + c_{n_s} y_{n_s} + \sum_{l_s \in L_s} \bar{c}_{l_s} u_{l_s} + c_{l_s} y_{l_s} \tag{1a}$$

$$\text{s.t.} \sum_{n_s \in N_s} \varphi_{n_s}^{n_v} x_{n_s}^{t,n_v} = 1 \qquad \forall t \in T, n_v \in N_v(t) \tag{1b}$$

$$\sum_{t \in T} d_t x_{n_s}^{t,n_v} \leq \kappa_{n_v} y_{n_s}^{n_v} \qquad \forall n_v \in N_v, n_s \in N_s \tag{1c}$$

$$\sum_{n_v \in N_v} \kappa_{n_v} y_{n_s}^{n_v} \leq u_{n_s} \qquad \forall n_s \in N_s \tag{1d}$$

$$\sum_{t \in T} \sum_{l_v \in L_v(t)} d_t f_{l_s}^{t,l_v} \leq u_{l_s} \qquad \forall l_s \in L_s \tag{1e}$$

$$u_{n_s} \leq \bar{\kappa}_{n_s} + \kappa_{n_s} y_{n_s} \qquad \forall n_s \in N_s \tag{1f}$$

$$u_{l_s} \leq \bar{\kappa_{l_s}} + \kappa_{l_s} y_{l_s} \qquad \forall l_s \in L_s \tag{1g}$$

$$\sum_{j \in N_s(n_s)} (f_{(n_s,j)}^{t,l_v} - f_{(j,n_s)}^{t,l_v}) = x_{n_s}^{t,n_{v_1}} - x_{n_s}^{t,n_{v_2}} \qquad \forall t \in T, l_v \in L_v(t), n_s \in N_s \tag{1h}$$

$$x_{n_s}^{t,n_v}, f_{l_s}^{t,l_v} \in \mathbb{R}_{[0,1]}, y_{n_s}^{n_v}, y_{n_s}, y_{l_s} \in \mathbb{Z}_{\geq 0}, \quad u_{n_s}, u_{l_s} \in \mathbb{R}_{\geq 0} \tag{1i}$$

where $f_{l_s}^{t,l_v} := f_{(n_{s_1},n_{s_2})}^{t,l_v} + f_{(n_{s_2},n_{s_1})}^{t,l_v}$ for all $l_s \in L_s$. Objective function (1a) minimises the combined costs of substrate resource utilisation and potential capacity installation when hosting the network slice. Constraints (1b) ensure that demand t is routed in its entirety through an ordered set of virtual functions $N_v(t)$ hosted on suitable substrate nodes. Constraints (1c) denote the capacity requirements of the virtual functions. Constraints (1d) guarantee that the capacities reserved at the substrate nodes are higher than the capacity requirements of the virtual functions. Analogously, constraints (1e) ensure the capacities reserved at the substrate links are higher than the volume of traffic routed through them. Constraints (1f) resp. (1g) impose that the reserved capacities at the substrate nodes resp. links do not exceed the total available capacities (i.e., the residual capacities and the additionally installed capacity modules). Constraints (1h) conserve flow at the substrate nodes.

III. Robust Network Slice Design Models

We previously introduced the deterministic version of the general network slice design problem. This is appropriate under the assumption that the forecast traffic demands are accurate. In practice, however, the task of accurately estimating the future traffic behaviour is challenging due to the dynamic nature of next generation network services [4]. This clearly poses a problem for the operation of network slices, especially for services supporting lifeline communications where maintaining service continuity is of paramount importance. It is thus imperative that the information regarding traffic uncertainty is incorporated into the model formulation to be able to cope with the stochastic nature of the traffic demands. Robust Optimisation (RO) has been increasingly adopted to address the challenges arising from parametric uncertainties. By choosing an appropriate uncertainty set to model the parametric uncertainties, RO makes it possible to obtain cost-effective solutions that are deterministically feasible for any realisation of the uncertain parameters within the contours of the defined uncertainty set. For an expansive treatment on RO, we refer the reader to [18]. Recent applications of RO in the field of network optimisation include [9], [10], [11].

In this work, we consider the widely adopted robust optimisation approach proposed by Bertsimas and Sim [1], [2] to model uncertainties in the traffic demands. The Γ-robust approach is favourable because of two reasons: First, it allows the feasibility level to be easily adjusted by tuning the parameter Γ, which corresponds to the number of simultaneous parametric deviations against which we intend to be protected. Secondly, the resulting reformulations of the robust counterparts are computationally tractable, thereby making it appealing from a practitioner's perspective.

A. The Robust Network Slice Design Problem: Uncorrelated Demand Uncertainties

We model the traffic demands as symmetric and independent random variables bound by the interval $[\bar{d}_t - \hat{d}_t, \bar{d}_t + \hat{d}_t]$, where $\bar{d}_t$ denotes the nominal value of the forecast traffic volume, and $\hat{d}_t$, in turn, denotes the maximum deviation from the forecast traffic volume. Assuming at most $\Gamma \in \mathbb{Z}^{|T|}$ demands simultaneously attain their worst-case realisations, we can now replace the capacity constraints (1c) and (1e) with the following non-linear robust counterparts:

$$\sum_{t \in T} \bar{d}_t x_{n_s}^{t,n_v} + \max_{\substack{D \subseteq T: \\ |D| \leq T}} \sum_{t \in D} \hat{d}_t x_{n_s}^{t,n_v} \leq \kappa_{n_v} y_{n_s}^{n_v} \quad \forall n_v \in N_v, n_s \in N_s \quad \text{(1c')}$$

$$\sum_{t \in T} \sum_{l_v \in L_v(t)} \bar{d}_t f_{l_s}^{t,l_v} + \max_{\substack{D \subseteq T: \\ |D| \leq T}} \sum_{t \in D} \sum_{l_v \in L_v(t)} \hat{d}_t f_{l_s}^{t,l_v} \leq u_{l_s} \quad \forall l_s \in L_s \quad \text{(1e')}$$

The non-linear robust capacity constraints can be linearised by casting them as exponential many linear constraints, each accounting for a possible realisation of the uncertain demand scenario. Thus, the exponential model for the Γ-robust network slice design problem for uncorrelated demand uncertainties takes the form:

(1a), (1b), (1d), (1f), (1g), (1h), (1i)

$$\sum_{t \in T} \bar{d}_t x_{n_s}^{t,n_v} + \sum_{t \in D} \hat{d}_t x_{n_s}^{t,n_v} \leq \kappa_{n_v} y_{n_s}^{n_v} \quad \forall n_v \in N_v, n_s \in N_s \quad \text{(2a)}$$

$$\sum_{t \in T} \sum_{l_v \in L_v(t)} \bar{d}_t f_{l_s}^{t,l_v} + \sum_{t \in D} \sum_{l_v \in L_v(t)} \hat{d}_t f_{l_s}^{t,l_v} \leq u_{l_s} \quad \forall l_s \in L_s \quad \text{(2b)}$$

where constraints (2a), (2b) are applicable for every demand subset $D \subseteq T$ provided $|D| \leq \Gamma$. The problem of having to include an exponential set of constraints in the optimisation model (2) can be circumvented as follows: For a solution (x, f) to problem (1), we compute the violated inequalities of capacity constraints (2a), (2b) for the obtained solution by reverting to separation procedures elaborated in the works of [9], [19]. We subsequently add these violated inequalities to model (1) and repeat the procedure until no more violations are observed.

An alternate approach to the above method is to employ the dualisation technique proposed by Bertsimas and Sim [1], [2] in order to obtain a tractable reformulation of the exponential Γ-robust NSDP. By virtue of LP duality, the inner maximisation problem in constraint (1c') for a fixed $x_{n_s}^{t,n_v}$ can be replaced with its corresponding dual equivalent:

$$\max_{\substack{D \subseteq T: \\ |D| \leq \Gamma}} \sum_{t \in D} \hat{d}_t x_{n_s}^{t,n_v} = \min \sum_{t \in T} \rho_{n_s}^{t,n_v} + \Gamma \pi_{n_v}^{n_v}$$
$$\text{s.t.} \quad \rho_{n_s}^{t,n_v} + \pi_{n_s}^{n_v} \geq \hat{d}_t x_{n_s}^{t,n_v} \quad \forall t \in T$$
$$\rho_{n_s}^{t,n_v}, \pi_{n_s}^{n_v} \in \mathbb{R}_{\geq 0} \quad \forall t \in T$$

where $\rho_{n_s}^{t,n_v}, \pi_{n_s}^{n_v}$ are the dual variables. The dual equivalent of the inner maximisation problem in constraint (1e') can be obtained in a similar manner.

The compact reformulation of the Γ-robust network slice design problem for uncorrelated demand uncertainties reads:

(1a), (1b), (1d), (1f), (1g), (1h), (1i)

$$\sum_{t \in T} \bar{d}_t x_{n_s}^{t,n_v} + \sum_{t \in T} \rho_{n_s}^{t,n_v} + \Gamma \pi_{n_s}^{n_v} \leq \kappa_{n_v} y_{n_s}^{n_v} \quad \forall n_v \in N_v, n_s \in N_s \quad \text{(3a)}$$

$$\rho_{n_s}^{t,n_v} + \pi_{n_s}^{n_v} \geq \hat{d}_t x_{n_s}^{t,n_v} \quad \forall t \in T, n_v \in N_v(t), n_s \in N_s \quad \text{(3b)}$$

$$\sum_{t \in T} \sum_{l_v \in L_v(t)} d_t f_{l_s}^{t,l_v} + \sum_{t \in T} \rho_{l_s}^{t} + \Gamma \pi_{l_s} \leq u_{l_s} \quad \forall l_s \in L_s \quad \text{(3c)}$$

$$\rho_{l_s}^t + \Gamma\pi_{l_s} \geq \sum_{l_v \in L_v(t)} \hat{d}_t f_{l_s}^{t,l_v} \qquad \forall t \in T, l_s \in L_s \quad (3d)$$

$$\rho_{n_s}^{t,n_v}, \pi_{n_s}^{n_v}, \rho_{l_s}^t, \pi_{l_s} \in \mathbb{R}_{\geq 0} \quad (3e)$$

Note that the spatial correlations between the uncertain traffic demands are not explicitly included in the definition of the uncertainty set, and hence we term the model *uncorrelated*. This may however have consequences on the robustness of the network slice design solutions if the uncertain traffic demands exhibit strong correlation, as we will show in Section IV-B.

B. The Robust Network Slice Design Problem: Correlated Demand Uncertainties

Recall that, in Section III-A, we modelled the demands as independent and symmetrically distributed random variables in the pre-defined interval $[\bar{d}_t - \hat{d}_t, \bar{d}_t + \hat{d}_t]$. We now consider the scenario where the uncertain traffic demands are no longer independent, but possibly correlated. This reflects the case of lifeline communications, where the unforeseen events occurring at certain locations clearly have an impact on the independence of the uncertain traffic demands. Such a scenario potentially resembles a flash-crowd or a traffic surge where some spatial correlation exists between different traffic demands. Hence, our previous assumption of independence among the uncertain traffic demands no longer holds true and this requires us to re-define our uncertainty set so as to explicitly include the correlation between the demands. Bertsimas and Sim [2] propose an alternate model to handle correlated parametric uncertainties, which seeks to identify a subset of uncertain sources of cardinality $\Gamma \in \mathbb{Z}^{|K|}$ that have the worst impact on each parameter. On the contrary, we intend to be protected against a subset of demands that assume their worst-case realisations due to the impact of a set K of uncertain sources. As a result, we apply refinements to the definition of the uncertainty set as outlined in the work of Gregory et al [12].

We now characterise the uncertainty set for correlated traffic demands: The uncertain traffic demands are modelled as symmetric and independent random variables bound by $[\bar{d}_t - \sum_{k\in K} g_{k,t}, \bar{d}_t + \sum_{k\in K} g_{k,t}]$, where $\bar{d}_t$ corresponds to the nominal value of the forecast traffic volume, and $g_{k,t}$ refers to the impact of the uncertainty source k on demand t. Note that by explicitly incorporating the impact of the uncertainty sources on each demand t, we guarantee that the correlation among the traffic demands is now accounted for in the subsequent robust formulations. We assume at most $\Gamma \in \mathbb{Z}^{|T|}$ demands deviate from their nominal value d_t simultaneously. Hence, we replace the nominal constraints (1c), (1e) with their respective robust counterparts:

$$\sum_{t\in T} \bar{d}_t x_{n_s}^{t,n_v} + \max_{\substack{D\subseteq T \\ |D|\leq\Gamma}} \sum_{t\in D}\sum_{k\in K} g_{k,t} x_{n_s}^{t,n_v} \leq \kappa_{n_v} y_{n_s}^{n_v}$$
$$\forall n_v \in N_v, n_s \in N_s \quad (1c")$$

$$\sum_{t\in T}\sum_{l_v\in L_v(t)} \bar{d}_t f_{l_s}^{t,l_v} + \max_{\substack{D\subseteq T \\ |D|\leq\Gamma}} \sum_{t\in D}\sum_{k\in K}\sum_{l_v\in L_v(t)} g_{k,t} f_{l_s}^{t,l_v} \leq u_{l_s}$$
$$\forall l_s \in L_s \quad (1e")$$

Analogous to the previous section, we can replace the non-linear robust capacity constraints (1c"), (1e") with an exponential family of linear inequalities, each corresponding to a possible realisation of the uncertain demand scenario. We now present the exponential model for the Γ-robust network slice design problem for correlated demand uncertainties:

(1a), (1b), (1d), (1f), (1g), (1h), (1i)

$$\sum_{t\in T} \bar{d}_t x_{n_s}^{t,n_v} + \sum_{t\in D}\sum_{k\in K} g_{k,t} x_{n_s}^{t,n_v} \leq \kappa_{n_v} y_{n_s}^{n_v}$$
$$\forall n_v \in N_v, n_s \in N_s \quad (4a)$$

$$\sum_{t\in T}\sum_{l_v\in L_v(t)} \bar{d}_t f_{l_s}^{t,l_v} + \sum_{t\in D}\sum_{k\in K} g_{k,t} f_{l_s}^{t,l_v} \leq u_{l_s}$$
$$\forall l_s \in L_s \quad (4b)$$

where constraints (4a), (4b) hold true for every subset of traffic demands $D \subseteq T$ and $|D| \leq T$. Notwithstanding the presence of exponential many constraints, we can employ separation oracles to solve the exponential model for the Γ-robust NSDP for correlated demand uncertainties.

Exploiting LP duality, and fixing $x_{n_s}^{t,n_v}$, the inner maximisation problem in (1c") can be linearised as follows:

$$\max_{\substack{D\subseteq T: \\ |D|\leq\Gamma}} \sum_{t\in D}\sum_{k\in K} g_{k,t} x_{n_s}^{t,n_v} = \min \sum_{t\in T} \rho_{n_s}^{t,n_v} + \Gamma\pi_{n_s}^{n_v}$$
$$\text{s.t. } \rho_{n_s}^{t,n_v} + \pi_{n_s}^{n_v} \geq \sum_{k\in K} g_{k,t} x_{n_s}^{t,n_v} \qquad \forall t \in T$$
$$\rho_{n_s}^{t,n_v}, \pi_{n_s}^{n_v} \in \mathbb{R}_{\geq 0} \qquad \forall t \in T$$

where $\rho_{n_s}^{t,n_v}$, $\pi_{n_s}^{n_v}$ are the dual variables. Constraints (1e") can be linearised in a similar fashion.

The exponential Γ-robust NSDP (4) can now be reformulated as the compact Γ-robust network slice design problem for correlated demand uncertainties:

(1a), (1b), (1d), (1f), (1g), (1h), (1i)

$$\sum_{t\in T} \bar{d}_t x_{n_s}^{t,n_v} + \sum_{t\in T} \rho_{n_s}^{t,n_v} + \Gamma\pi_{n_s}^{n_v} \leq \kappa_{n_v} y_{n_s}^{n_v}$$
$$\forall n_v \in N_v, n_s \in N_s \quad (5a)$$

$$\rho_{n_s}^{t,n_v} + \pi_{n_s}^{n_v} \geq \sum_{k\in K} g_{k,t} x_{n_s}^{t,n_v}$$
$$\forall t \in T, n_v \in N_v(t), n_s \in N_s \quad (5b)$$

$$\sum_{t\in T}\sum_{l_v\in L_v(t)} \bar{d}_t f_{l_s}^{t,l_v} + \sum_{t\in T} \rho_{l_s}^t + \Gamma\pi_{l_s} \leq u_{l_s} \qquad \forall l_s \in L_s \quad (5c)$$

$$\rho_{l_s}^t + \pi_{l_s} \geq \sum_{k\in K}\sum_{l_v\in L_v(t)} g_{k,t} f_{l_s}^{t,l_v} \qquad \forall t \in T, l_s \in L_s \quad (5d)$$

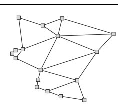
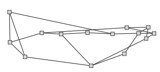
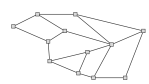
$$\rho_{n_s}^{t,n_v}, \pi_{n_s}^{n_v}, \rho_{l_s}^t, \pi_{l_s} \in \mathbb{R}_{\geq 0} \quad (5e)$$

IV. Performance Evaluation

In this section, we analyse the performance of the proposed robust models (2)-(5) in terms of their solution quality. We then evaluate the relative costs and the robustness of the solutions obtained from these models, and investigate the benefits of explicitly considering the spatial correlations within the uncertainty set construction.

Instances: We construct problem instances similar to that of [15] while suitably modifying them to account for spatially correlated traffic uncertainties. For the physical substrate network, we consider three sample network topologies from SNDlib [16]: NOBEL-DE $(17, 26)$, NOBEL-US $(14, 21)$, POLSKA $(12, 18)$. The spare physical substrate network resources (node/link) are randomly drawn from the value set $(300.0, 400.0, 500.0)$ weighted by $(0.3, 0.4, 0.3)$ while the cost per occupied unit of these resources is set to 2.5 units. The size and cost of a capacity module is set to 250 units. The cost to install a capacity module on either a substrate node or link is set to 250 units as well.

TABLE I
SUBSTRATE NETWORK EXAMPLES

Network	NOBEL-DE	NOBEL-US	POLSKA
Topology			
$\|\boldsymbol{N_s}\|$	17	14	12
$\|\boldsymbol{L_s}\|$	26	21	18
$\|\boldsymbol{T}\|$	34	28	24

We consider the use case of a generic network slice design where the set of virtual functions is given as N_v ={VF1, VF2, VF3, VF4, VF5} while the sequence in which the end-to-end traffic flows traverse the virtual functions is expressed as: L_v ={($i(t)$, VF1), (VF1, VF2), (VF2, VF3), (VF3, VF4), (VF4, VF5), (VF5, $e(t)$)}. As for the set of traffic demands T, we generate $2 \cdot |N_s|$ demands whose source nodes $i(t)$ and sink nodes $e(t)$ are drawn from the set N_s of physical substrate nodes uniformly at random while ensuring a non-zero hop count. For our evaluation, it is assumed that all substrate network nodes have the capability to host a virtual function of any type $n_v \in N_v$. The resource allocation granularity κ_{n_v} for all virtual functions $n_v \in N_v$ is set to 100 units.

Due to the unavailability of spatially correlated traffic datasets, we generate synthetic traffic traces by applying the following four-step procedure: First, for each demand $t \in T$, we randomly draw a value from $(20.0, 30.0, 40.0)$ with probability $(0.3, 0.4, 0.3)$. Second, to ensure that the traffic coefficients d_t are normally distributed, for each $t \in T$, we draw 1440 samples at random from a normal distribution with mean 0 and standard deviation of 50% of the respective value chosen in the first step. At this stage, a quick verification of the traffic traces reveal - as expected - negligible correlations. To enforce spatial correlation among the demands, the samples generated in the second step are scaled by the Cholesky factorisation of a manually-built correlation matrix such that $\varrho_{t,t'} = 0.99$ for all $t, t' \in T : t \neq t'$. Finally, for every $t \in T$, the value determined in the first step is added to each of these 1440 samples to obtain spatially correlated synthetic traffic snapshots. We now determine the nominal $(\bar{d}_t)$ and uncertain $(\hat{d}_t)$ volume of each demand as the arithmetic mean and 3σ of the generated traffic traces, respectively. For the case of the correlated models (4), (5), however, the task of precisely identifying the set K of uncertain sources and their impact $g_{k,t}$ on the uncertain traffic demands is indeed challenging. Alternatively, we use the correlations among the traffic demands under the assumption that the covariance matrix Σ of the uncertain traffic demands is available [20]. We then apply Cholesky decomposition to obtain the lower triangular matrix $\boldsymbol{L}$, wherein each element $L_{t,t'}$ captures the impact of the source of traffic uncertainty t' on the demand t which is suitably scaled by a factor of 3.

Setting: We implemented compact reformulations (3) and (5) in JuMP v0.18 [21] - a modelling language for mathematical optimisation embedded in Julia v0.6 [22] using the Gurobi Optimizer v7.0.0 [23] as the underlying MIP solver. For the exponential formulations (2) and (4), the violated robust inequalities are separated on-the-fly as lazy constraints which are then added to the formulations using the solver callback functionality supported by JuMP.

The computations are performed on a Linux machine with Intel® Core™ i7-4770K CPU @ 3.40 GHz and 32 GB RAM. We set a time limit of 3600 seconds for solving each problem instance. For each of the considered network topologies, models (2), (3), (4) and (5) are solved for $\Gamma = 0, \cdots, 15$ leading to a total of 192 network slice design problem instances.

A. *Separation vs. Reformulation*

To study the computational performance of the proposed solution methodologies - separation method (for models (2), (4)) and compact reformulation (for models (3), (5)) - we compare the solutions obtained from the respective methodologies in terms of their optimality gaps. The optimality gap is defined as the relative gap between the best integer objective and the best lower bound attained by the solver and is often used as a qualitative indicator of the obtained solution. A positive optimality gap indicates the distance of the obtained solution from optimality whereas a solution with an optimality gap of 0 is interpreted as the optimal solution.

Figure 1 depicts the optimality gaps of the solutions obtained from the separation method against those of compact reformulation for both the uncorrelated and correlated models. Every mark in the figure corresponds to a robust network slice design for a specific configuration of (nominal, uncertain, Γ). With the increment in the protection level Γ, the resulting problem instances become harder to solve leading to higher optimality gaps (which are consequently represented by darker marks). The marks lying on the bisecting line, however, represent solutions of the same quality irrespective of the adopted solution methodology.

For each network topology, we observe a higher presence of the marks to the left of the bisecting line indicating that the solutions obtained from the separation method are of inferior quality compared to the compact reformulation method. The reason for this behaviour is mainly attributed to the time spent by the separation routines to identify violated robust

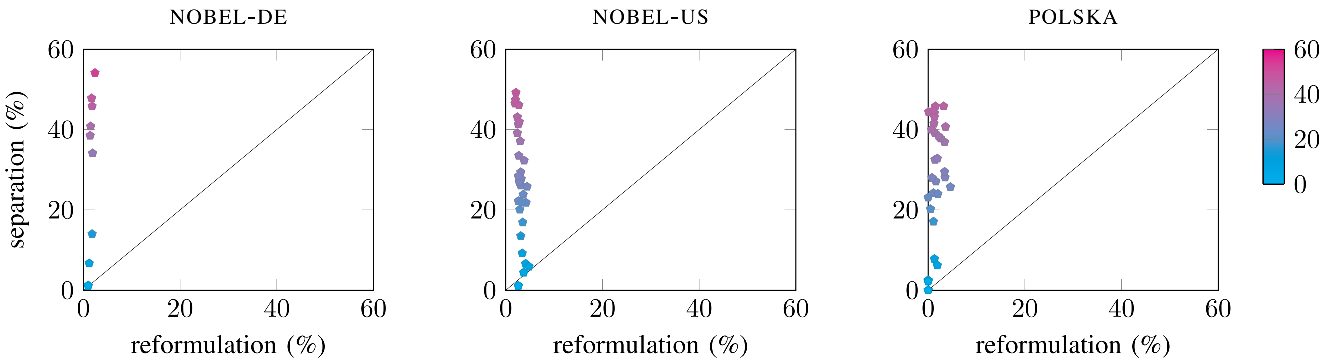


Fig. 1. Comparison of optimality gaps of the solutions obtained via separation and reformulation methods for all problem instances.

inequalities at the active nodes of the branch-and-bound tree. In particular, of the considered 96 problem instances for each of the methods, the compact reformulation always returned a solution with an average optimality gap of 2.17%. On the contrary, the separation method was not able to solve 23 of the considered problems within the imposed time limit, notably for the NOBEL-DE problem instance for the case of both the uncorrelated and correlated models. This behaviour is manifested by the sparsely distributed markers for the problem instance NOBEL-DE. Interestingly, we did not notice a considerable difference ($< 1.6\%$) in the optimality gaps of the solutions obtained for both the uncorrelated and correlated models irrespective of the solution method implying that the inclusion of correlation within the uncertainty set doesn't increase the problem complexity.

To summarise, the compact reformulation proves to be a computationally viable alternative to the separation method, yielding solutions of lower optimality gap for most of the problem instances (in our case 92 of the considered 96 problem instances) whereas the separation procedure fails to return a non-trivial solution for 23 instances. Henceforth, we employ the compact reformulation method to evaluate the performance of the (uncorrelated and correlated) robust NSDP models in the subsequent section.

B. Cost vs. Realised Robustness

In our second study, we evaluate the performance of the proposed robust network slice design models in terms of the relative costs and the realised robustness of the obtained solutions. Consistent with the earlier works of [8], [9], we define price of robustness as the percentual increase in the cost of the solution to provide robustness guarantees while the realised robustness is defined as the percentage of snapshots/traffic matrices for which the obtained solution is still feasible.

In Figure 2, we illustrate the results obtained for the NOBEL-US problem instance from models (3) and (5). Every mark represents the relative cost and the robustness of the obtained network slice design solution for a particular (nominal, uncertain, Γ) configuration. As anticipated, when increasing the protection factor Γ, the robustness of the solutions improve, and as a result, the cost to provide robustness increases. While this trend remains irrespective of the considered robust model, we notice that the conservativeness (i.e. the cost vs. robustness tradeoff) of the solutions obtained from both models are quite similar. This implies that despite the inclusion of correlation into the uncertainty set, the solutions from the correlated model do not add significantly to the conservativeness. However, an interesting observation is that for a fixed (nominal, uncertain, Γ) setting, the correlated model yields solutions that are more robust compared to the uncorrelated model. Encouraged by this observation, in the following, we investigate in further detail, the robustness of the solutions obtained from the uncorrelated and correlated models.

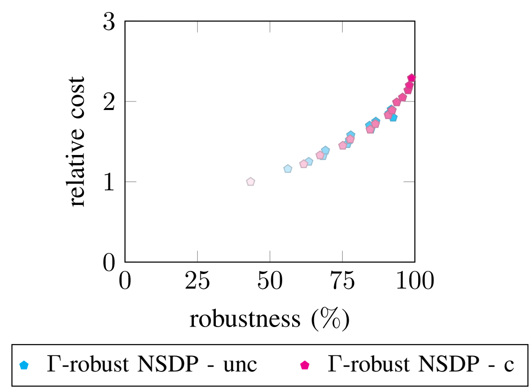


Fig. 2. Relative cost (w.r.t. the non-robust model (1)) vs. realised robustness for NOBEL-US.

Figure 3 visualises the realised robustness of the obtained solutions for both the uncorrelated and correlated network slice design models. Each mark in the figure indicates the robustness of the network slice design solution obtained for the uncorrelated and correlated models for a particular (nominal, uncertain, Γ) setting. As mentioned earlier, we observe that for every increase in the protection factor Γ, the robustness of the obtained solution improves which is reflected by the color of the marks. For all networks, except for $\Gamma = 0$ whose mark is located on the bisecting line, the marks are positioned above the bisecting line indicating that the network slice design solutions obtained from the correlated model provide a higher degree of robustness (at most 9 percentage points) compared to the uncorrelated model. This is an expected yet predominant

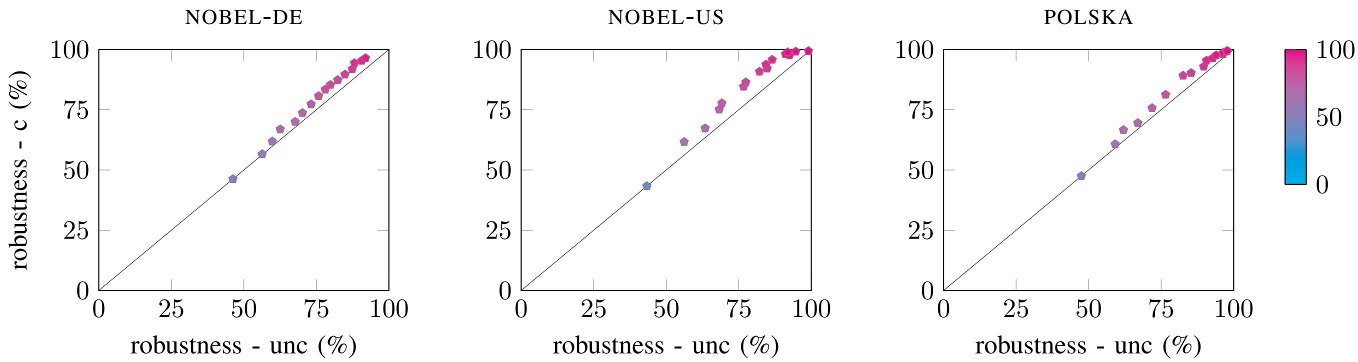


Fig. 3. Realised robustness (correlated) vs. realised robustness (uncorrelated) for all problem instances.

consequence of including the correlation information in the construction of the uncertainty set as outlined in Section III.

V. Conclusions

In this paper, we consider the network slice design problem outlined in [15]. We suitably extend its formulation to account for i) traffic demand uncertainties by adopting the Γ-robust uncertainty set of Bertsimas and Sim [1], [2], and ii) correlation among the uncertain traffic demands by explicitly including the correlation information in the uncertainty set construction. We then present two solution approaches - separation method and compact reformulation - to solve the robust network slice design models with and without traffic demand correlation.

Our computational study shows that the compact reformulation outperforms the separation method, yielding solutions of lower optimality gaps for the majority of the considered problem instances. In a further step, we evaluate the robustness of the solutions obtained from the uncorrelated and correlated robust NSDP models. As an outcome of explicitly integrating the information of traffic demand correlations into the uncertainty set, the correlated model is able to yield solutions that are more robust against uncertain traffic demands while maintaining a similar level of conservativeness. In future, we plan to investigate an alternative method based on the histogram model to characterise the correlation in the uncertain traffic demands.

References

[1] D. Bertsimas and M. Sim, "Robust discrete optimization and network flows," *Mathematical Programming*, vol. 98, no. 1, pp. 49–71, 2003.

[2] ——, "The Price of Robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.

[3] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021," Tech. Rep., Feb 2017.

[4] NGMN Alliance, "5G White Paper," Tech. Rep., Feb 2015.

[5] M. Richart, J. Baliosian, J. Serrat, and J. L. Gorricho, "Resource Slicing in Virtual Wireless Networks: A Survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 462–476, Sept 2016.

[6] F. Esposito, I. Matta, and V. Ishakian, "Slice Embedding Solutions for Distributed Service Architectures," *ACM Comput. Surv.*, vol. 46, no. 1, pp. 6:1–6:29, Jul. 2013.

[7] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, "Virtual Network Embedding: A Survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1888–1906, Fourth 2013.

[8] S. Coniglio, A. M. C. A. Koster, and M. Tieves, "Data Uncertainty in Virtual Network Embedding: Robust Optimization and Protection Levels," *Journal of Network and Systems Management*, vol. 24, no. 3, pp. 681–710, 2016.

[9] A. M. C. A. Koster, M. Kutschka, and C. Raack, "Robust Network Design: Formulations, Valid Inequalities, and Computations," *Networks*, vol. 61, no. 2, pp. 128–149, 2013.

[10] T. Bauschert, C. Büsing, F. D'Andreagiovanni, A. M. C. A. Koster, M. Kutschka, and U. Steglich, "Network Planning under Demand Uncertainty with Robust Optimization," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 178–185, February 2014.

[11] F. D'Andreagiovanni, J. Krolikowski, and J. Pulaj, "A fast hybrid primal heuristic for multiband robust capacitated network design with multiple time periods," *Applied Soft Computing*, vol. 26, no. Supplement C, pp. 497 – 507, 2015.

[12] C. Gregory, K. Darby-Dowman, and G. Mitra, "Robust optimization and portfolio selection: The cost of robustness," *European Journal of Operational Research*, vol. 212, no. 2, pp. 417 – 428, 2011.

[13] R. Ferreira, L. Barroso, and M. Carvalho, "Demand response models with correlated price data: A robust optimization approach," *Applied Energy*, vol. 96, pp. 133 – 149, 2012.

[14] A. Moreira, A. Street, and J. M. Arroyo, "Energy and reserve scheduling under correlated nodal demand uncertainty: An adjustable robust optimization approach," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 91 – 98, 2015.

[15] A. Baumgartner, T. Bauschert, A. M. C. A. Koster, and V. S. Reddy, "Optimisation Models for Robust and Survivable Network Slice Design: A Comparative Analysis," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–7.

[16] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, "SNDlib 1.0-Survivable Network Design Library," *Networks*, vol. 55, no. 3, pp. 276–286, 2010.

[17] M. Barcelo, J. Llorca, A. M. Tulino, and N. Raman, "The Cloud Service Distribution Problem in Distributed Cloud Networks," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 344–350.

[18] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, ser. Princeton series in applied mathematics. Princeton, NJ: Princeton Univ. Press, 2009.

[19] D. Bertsimas, I. Dunning, and M. Lubin, "Reformulation versus cutting-planes for robust optimization," *Computational Management Science*, vol. 13, no. 2, pp. 195–217, 2016.

[20] A. Jalilvand-Nejad, R. Shafaei, and H. Shahriari, "Robust optimization under correlated polyhedral uncertainty set," *Computers & Industrial Engineering*, vol. 92, pp. 82 – 94, 2016.

[21] I. Dunning, J. Huchette, and M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[22] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.

[23] I. Gurobi Optimization, "Gurobi Optimizer Reference Manual," 2016. [Online]. Available: http://www.gurobi.com