How physical network topologies affect virtual network embedding quality: A characterization study based on ISP and datacenter networks

Marcelo Caggiani Luizelli *, Leonardo Richter Bays, Luciana Salete Buriol, Marinho Pilla Barcellos, Luciano Paschoal Gaspery

Federal University of Rio Grande do Sul, Av. Bento Gonçalves, 9500, Campus do Vale – Bloco IV, 91501-970 Porto Alegre, Brazil

ABSTRACT

Network virtualization is a mechanism that allows the coexistence of multiple virtual networks on top of a single physical substrate. Due to its well-known potential benefits (e.g., lower CAPEX/OPEX expenditures), it has been embraced by the IT sector, specially by Internet Service Providers (ISPs) and cloud computing/datacenter companies. One of the research challenges addressed recently in the literature is the efficient mapping of virtual resources on physical infrastructures. Although this challenge has received considerable attention, state-of-the-art approaches present, in general, a high rejection rate, i.e., the ratio between the number of denied virtual network requests and the total amount of requests is considerably high. In this work, we investigate the relationship between the quality of virtual network mappings and the topological structures of the underlying substrates. Exact solutions of an online embedding model are evaluated under different classes of ISP and datacenter network topologies. The obtained results demonstrate that the employment of physical topologies that contain regions with high connectivity significantly contributes to the reduction of rejection rates and, therefore, to improved resource usage. Additionally, through an extensive analysis of denied requests, we assess the main rejection causes related to both ISP and datacenter networks and provide strong evidence of each one. In summary, through the embedding of virtual requests, available resources in ISP networks tend to be more partitioned in comparison to datacenter networks. Such differences on partitioning levels lead to a different percentage of rejection causes in each topology class.

© 2016 Elsevier Ltd. All rights reserved.
1. Introduction

Network virtualization is a mechanism that allows the coexistence of multiple heterogeneous virtual networks (VNs) sharing resources of the same physical substrate. The architectures, protocols, and topologies used in these VNs are unconstrained by the substrate network on which they are instantiated. Through network virtualization, Infrastructure Providers (InPs) are able to easily allocate and deallocate virtual networks with proper resource isolation. In other words, this mechanism enables InPs to support the creation of custom networks on demand, meeting different requirements imposed by requesters. ISP and datacenter providers have been taking advantage of network virtualization in order to cope with limitations of such architectures, e.g., performance isolation, limited management flexibility, and security risks (Zhani et al., 2013; Bays et al., 2015).

One of the major research challenges in network virtualization is the efficient mapping of physical resources to virtual networks (VNE – Virtual Network Embedding). The resource mapping process must consider the capacities of physical network devices, as well as the demands of virtual networks (for instance, virtual link bandwidth and processing capacity of virtual routers). Although previous work explores the problem of online virtual network embedding in the context of Internet Service Providers (ISPs) (Yu et al., 2008; Cheng et al., 2011; Chowdhury et al., 2012; Alkmim et al., 2013; Richter Bays et al., 2014; Oliveira et al., 2015; Beck et al., 2015; Zhang et al., 2015) and datacenter networks (Guo et al., 2010; Ballani et al., 2011; Stefani Marcon et al., 2013; Zhang et al., 2014), considerably high rejection rates\(^1\) are commonly observed (as high as 53%). We assume that a subset of these rejections is caused by temporary resource exhaustion, i.e., periods in which the available resources in the infrastructure as a whole are not sufficient to meet the demand. We theorize, however, that most rejections occur in situations in which a significant amount of resources is available, but a few saturated devices and links, depending on connectivity features of the physical substrate, hinder the acceptance of new requests.

Fig. 1 shows an example of a common cause for rejection of virtual requests. In it, two physical network substrates with different topological characteristics are depicted. Physical substrate 1 has better path diversity as well as a cycle connecting the required endpoints. Substrate 2, in contrast, has a single bridge link connecting the required endpoints. For the sake of simplicity, consider two identical virtual network requests with exactly one virtual link connecting its endpoints. Moreover, consider that each physical link can host at most one virtual link. The different topological features of Substrate B make it impossible to embed the second virtual network request on top of it, despite the presence of idle resources in other parts of the topology. The suggested mappings represented in this figure evidence that differences on how physical devices are interconnected may dramatically affect the ability to embed new requests (in this case by a factor of 2).

Despite efforts to solve the virtual network embedding problem, little has been done in order to extensively investigate the influence of different classes of network topologies in the process of virtual network embedding. Moreover, previous work in this area has considered topologies that often do not reflect those observed in commercial networks (Haddadi et al., 2008). Understanding the relationship between the employment of different network topologies and the mapping process is important to determine how certain topological features influence this process. For example, topologies with higher connectivity in strategic regions may favor a better utilization of physical resources, which in turn may lead to lower rejection rates. Such outcomes have the potential to raise the profit obtained by InPs and, at the same time, reduce costs for virtual network requesters.

In comparison to our previous seminal work on this subject (Caggiani Luizelli et al., 2013), in this paper we present an extended, revised characterization study in which we conduct an in-depth analysis of the impact of different classes of topologies typically employed in commercial infrastructures. More specifically, we formalize an optimal virtual network embedding model and evaluate it on substrates with different types of ISP (star, ladder, and hub & spoke) and datacenter (conventional multi-rooted tree, Bari et al., 2013, Fat-Tree, Leiserson, 1985, and VL2, Greenberg et al., 2005) network topologies. We consider different metrics such as rejection rates and resource consumption for each analyzed topology. Additionally, we assess and provide strong evidence for the causes of rejection associated with each network topology. In summary, the main relevant contributions of this paper are two-fold: (i) the formalization of a general online embedding model which considers location constraints and is able to cope with both ISP and datacenter network infrastructures; and (ii) the extensive evaluation and discussion of the impact of different classes of network topologies in the embedding process.

The remainder of this paper is organized as follows. Section 2 presents a discussion of related work in the area of virtual network embedding, highlighting the topologies considered in each work and the rejection rates obtained. In Section 3 we characterize the types of network topologies considered in this work. In Section 4 we formalize the online virtual network embedding model. In Section 5 we present and evaluate the obtained results. Finally, in Section 6 we conclude this paper with final remarks.

2. Related work

In this section, we present related work in the area of virtual network embedding in the context of ISP and datacenter networks. We briefly summarize the proposed solutions, highlighting the types of physical and virtual topologies employed, as well as

---

\(^1\) The rejection rate is defined as the ratio between the number of rejected requests and the number of incoming ones. Virtual requests are rejected when the InP is not able to properly accommodate them on the physical infrastructure.
rejection rates obtained by embedding methods, when available. We begin by reviewing approaches in the context of ISP infrastructures and, afterwards, move on to approaches that focus on datacenter networks.

Yu et al. (2008) present an online virtual network embedding model supporting path splitting and migration. Path splitting improves the utilization of physical resources by embedding a higher number of virtual networks on the substrate, while the migration of virtual network elements aims to reoptimize physical resource usage. Router and link mapping are performed in distinct steps. The experiments employ randomly generated topologies for physical and virtual networks, with fixed connectivity of 50%. No rejection rates are presented.

Another online model, formulated by Chowdhury et al. (2012), also performs router and link mapping in distinct steps. However, location constraints are used to preselect physical routers on which virtual routers will be hosted. According to the authors, the preselection facilitates the subsequent stage of link mapping. This model also allows path splitting. Physical topologies used in the evaluation of this model have the same characteristics as those used by Yu et al. (2008) (randomly generated with fixed connectivity of 50%). Experiments are performed considering randomly generated physical topologies and with three types of VN topologies, namely random, hub & spoke, and full mesh. Rejection rates observed in experiments using random virtual topologies vary between 20% and 45%. Hub & spoke topologies led to a rejection rate of 35%, whereas the rejection rate of full mesh VNs varies between 40% and 45%.

Cheng et al. (2011, 2012) perform the mapping of virtual network elements by ranking routers and links according to their own capacity as well as the capacity of their neighbors. Virtual routers and links are mapped to similarly ranked physical devices and paths. According to the authors, this strategy aims to avoid the formation of bottlenecks on the physical network. In the evaluated scenarios, physical and virtual topologies are randomly generated and rejection rates vary between 15% and 25%.

Alkmim et al. (2013) propose a model that focuses on minimizing the time needed to transfer binary images of virtual routers (stored in repositories connected to the network) to the physical routers that will host them. Their model considers requirements related to router and link capacity, as well as location constraints. The evaluation scenarios employ organic topologies created using the BA-2 model (Albert and Barabási, 2000). On average, rejection rates in experiments performed by the authors are approximately 53%.

Richter Bays et al. (2014) introduce an online virtual network embedding model with security support (i.e., allows the provisioning of VNs with support to varying levels of security). The authors model the different overhead costs associated with security mechanisms (used to protect VNs) and propose a heuristic-based algorithm to find sub-optimal solutions in a timely fashion. Inline with previous related work, experiments are carried out considering organic model (Albert and Barabási, 2000) to model both virtual and physical topologies. Rejection rates range from 10% to 50% depending on the security level required.

In addition to approaches focusing on virtual network embedding in backbone networks, in recent years research efforts have been made towards the virtualization of datacenter networks. The main reason relies on the fact that these Datacenter networks do not provide performance guarantees to tenants (Bari et al., 2013; Ballani et al., 2011) and, consequently tenant's applications may be hurt by unpredictable network performance. Next, we briefly describe and discuss proposals that introduce the notion of employing virtualization in datacenter networks as a means to isolate tenant traffic and ultimately provide performance predictability to applications.

Guo et al. (2010) propose the Virtual Data Center (VDC) abstraction as the unit of resource allocation for multiple tenants. VDC is defined as a set of virtual machines (VM) with an associated service level agreement (SLA). Based on this abstraction, the authors introduce a centralized, greedy allocation algorithm for mapping virtual to physical machines with bandwidth guarantees. The algorithm attempts to allocate VDC requests to a set of nearby machines whenever possible so as to ensure the SLA required and, at the same time, reduce the provider's costs. In addition, the algorithm allows virtual resources (VMs and bandwidth) to be dynamically added to or removed from VDCs. Experiments conducted by the authors employ three types of physical topologies, namely VL2 (Greenberg et al., 2009), Fat-Tree (Leiserson, 1985) and BCube (Guo et al., 2009). Neither VN topologies nor rejection rates are mentioned in the paper.

Ballani et al. (2011) propose the implementation of two specific virtual network abstractions (or topologies) to ensure performance guarantees to tenants. The first, virtual cluster, connects all virtual machines to a single non-oversubscribed virtual switch; the second, virtual oversubscribed cluster, creates a two-tier infrastructure that is composed of sets of virtual clusters interconnected through a virtual root switch. The authors describe a greedy algorithm to allocate VNs into the physical network. The evaluation scenarios employ tree-based physical topologies and the two proposed abstractions as VN requests. Observed rejection rates vary between 5% and 20%.

In contrast to Guo et al. (2010) and Ballani et al. (2011), Stefani Marcon et al. (2013) propose a different embedding strategy which groups similar applications into the same virtual network. Grouping mutually trusted tenants into a single VN has the benefit of keeping the number of VNs (instantiated in the datacenter network) comparatively lower, therefore reducing the associated virtualization overheads. At the same time, such grouping has the potential to mitigate attacks (against privacy and integrity) of mutually untrusted tenants, as they will be hosted in distinct, isolated VNs. The proposed embedding strategy is divided into two sub-problems. The first maps mutual-trusted applications to virtual networks. The second sub-problem embeds each VN into the physical network. The authors evaluate the proposed strategy considering multi-rooted trees (Shieh et al., 2011) as physical infrastructures and tree-like topologies as VNs. Results show rejection rates varying from 20% to 55%.

Recently, Zhani et al. (2013) revisited the online VDC embedding problem and proposed a generic formulation. The model allows multiple types of physical/virtual resources (e.g., switches, storage, GPU, and fast memory hardware) to be managed and, therefore, is particularly suitable to embed virtual infrastructures on top of heterogeneous environments. The designed solution is based on a coordinated three-step heuristic which performs VM, link and switch mapping. The authors consider physical topologies following a VL2 (Greenberg et al., 2009) design. As for VN topologies, they employ both star and VL2. Rejection rates account for 15–20% of the incoming requests.

Zhang et al. (2014) propose a reliability and availability-aware VDC embedding model. The proposed model captures the characteristics of hardware failures and dependencies among virtual components (e.g., VMs and storage devices) to provide an availability-aware embedding framework. The experiments employ VL2 topologies for the physical networks and three topologies for VDC requests, namely, multi-tiered, partition-aggregate (which is the virtual cluster abstraction proposed by Ballani et al., 2011) and MapReduce. Rejection rates are not presented but, according to the authors, are proportional to the ones presented by Zhani et al. (2013).

As previously stated, little has been done in order to evaluate the results of virtual network embedding strategies considering...
different types of network topologies in a precise manner. Most investigations in this area employ topologies that may not faithfully represent the topological properties of infrastructure provider networks. Moreover, as previously stated, considerably high rejection rates are often observed. For these reasons, this study aims to understand how topologies that are typically employed in real physical substrates influence different aspects of virtual network embedding, such as the rejection of virtual network requests and physical resource usage.

3. Physical network topologies

We now review the characteristics and properties of the main relevant physical topologies employed by infrastructure providers (InPs). In the context of this paper, we assume, due to widespread use of network virtualization, network topologies employed by ISP and datacenter providers.

3.1. ISP network topologies

The most traditional ISP network topologies are known as ladder, star, and hub & spoke. Network topologies organized as ladder are characterized by the absence of hubs, i.e., nodes with high connectivity and concentration of flows. Additionally, the infrastructure is formed by a set of loops. This type of topology tends to have lower cost regarding the deployment of links (due to its low connectivity) and the distance between nodes (in terms of number of hops) is typically high. Star networks have a low number of hubs connected to numerous nodes which, in turn, have low connectivity. In this type of network, the distance between nodes tends to be low, but traffic tends to become concentrated on the hubs. The hub & spoke class is characterized by a comparatively higher number of hubs, which tend to be interconnected. Additionally, a high number of nodes is connected to one or more hubs. Fig. 2 illustrates examples of the three aforementioned topology classes.

Kamiyama et al. (2010) conducted a study that formalizes the classification of ISP networks into the three previously described topology classes. In their study, the authors analyzed 23 commercial backbone networks (publicly available) with sizes ranging from 21 to 128 nodes. Through this analysis, the authors defined a set of metrics that capture the main topological properties present in each infrastructure. Such metrics include, for example, the connectivity degree of the network and the presence of hub nodes. Thus, the authors map the relationship between these metrics and the type of network topology of the infrastructure, enabling the classification of ISP network topologies into one of the previously described classes.

In this paper, we consider these three classes as the basis for our investigation of the employment of ISP topologies in the embedding process. We consider this well accepted classification of ISP network topologies presented above and the systematic approach proposed by Kamiyama et al. (2010) to characterize and, ultimately, generate these topologies with high degree of fidelity (in relation to real networks).

3.2. Datacenter networks topologies

When it comes to datacenter network (DCN) topologies, certain operation requirements become more crucial, as is the case of scalability and fault tolerance. Scalability has to do with ensuring high throughput for customers of a multi-tenant facility and also allowing resources (server and network devices) to be scaled up and down, in a timely fashion, according to an elastic demand. Fault tolerance, in turn, is the ability to minimize service disruption due to failures of both computing and network resources. Since we are dealing with a shared substrate, even minor problems may compromise and affect several customers. As we shall confirm later in this paper, depending on the underlying network topology, these requirements may not be fully met. To illustrate how virtual network embedding quality (in terms of scalability and fault tolerance) is influenced by the substrate network, we consider both a conventional topology design and more
sophisticated, switch-oriented ones (VL2 and Fat-Tree topologies).

Conventional DCN topologies are widely used by current datacenters and are based on a three-tiered multi-rooted tree-like physical topology (Ballani et al., 2011; Benson et al., 2010). Fig. 3(a) shows an example of this type of topology. It consists of three layers: the access, the aggregation, and the core layer. The access layer is composed of the Top-of-Rack (ToR) switches, which connect servers mounted on every rack; the aggregation layer consists of devices that interconnect ToR switches in the access layer; and the core layer is formed by routers that interconnect switches in the aggregation layer. Additionally, every ToR switch may be connected to multiple aggregation switches for redundancy (usually 1 + 1 redundancy), and every aggregation switch is typically connected to multiple core switches.

In turn, switch-oriented topologies use only commodity switches to perform routing functions and follow a Clos-based design. Examples of such design include VL2 (Greenberg et al., 2009) and Fat-Tree (Leiserson, 1985) topologies, as shown in Figs. 3(b) and 3(c). VL2 is composed of multiple layers of switches in which each switch in a layer is connected to all switches in the upper and lower layers. This topology is suitable for large-scale datacenters and provides multiple uniform paths between servers, as well as a full bisection bandwidth.\footnote{The bisection bandwidth of the network is the worst-case bandwidth-wise segmentation of the network in two equally-sized partitions (Farrington et al., 2009; Curtis et al., 2012).} In turn, Fat-Tree topologies are usually organized in a non-oversubscribed $k$-ary tree-like structure, consisting of $k$-port switches. There are $k$ two-layer pods having exactly $k/2$ switches (aggregation switches) connected to the access layer. Each of the $(k/2)^2$ $k$-port core switches (in the figure, represented by the upper circles) has one port connected to each of $k$ pods (represented by the dotted squares).

Based on the ISP and datacenter topologies presented above, we analyze how the employment of those topological classes affects the VN mapping process. More specifically, we identify and characterize correlations among different topological characteristics, physical resource usage, and rejection rates.

4. Optimal model for VN embedding

As our objective was to capture the impact of the underlying substrate topology in the network embedding problem, we formulated it as an Integer Linear Programming problem and solved it exactly by using an optimization software. The employment of such model means that the results are the best possible one would obtain when considering a substrate topology or another. Next, we describe the inputs, variables, and constraints of this model. Superscript letters are used to represent whether a set or variable is related to virtual ($V$) or physical ($P$) resources, or whether it is associated to routers ($R$) or links ($L$).

Both physical topologies and virtual network requests are represented by directed graphs $\mathcal{N} = (R, L)$. Vertices $R$ represent routers, while each link $L$ represents a unidirectional link. Bidirectional links are represented as a pair of links in opposite directions (for instance, $(a,b)$ and $(b,a)$). Thus, the model allows the representation of any type of physical and virtual topology (Table 1).

Each physical router is associated with a location identifier stored in a set $S$. This enables virtual network requesters to indicate specific locations in which virtual routers must be instantiated (e.g., to ensure connectivity between two or more geographical locations). If a virtual router has a location requirement, it is stored in set $S^v$. Our approach of tagging routers with specific location identifiers differs from previous work (Chowdury et al., 2012; Fajjari et al., 2011), which treat locations simply as regions defined by radius.

In real life, physical routers have limited CPU and memory capacities. In our model such capacities are represented, respectively, by $C^P_i$ and $M^P_i$. Analogously, CPU and memory requirements of each virtual router on a network $r$ are represented by $C^v_{r,i}$ and $M^v_{r,i}$. Likewise, physical links have limited bandwidth capacity, represented by $B^P_{i,j}$, while the bandwidth required by each virtual link is represented by $B^v_{i,j}$.

The model takes as input virtual network requests and embeds them in an online manner. Thus, it is necessary to consider virtual elements that were previously embedded on the substrate. Previously embedded virtual routers are stored in set $E^v_{r,i}$, while previously embedded links are stored in set $E^L_{i,j,r,k}$.
Table 1
Glossary of symbols to the optimization model.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N^p = {R^p, L^p}$</td>
<td>Physical network infrastructure, composed of physical routers $R^p$ and links $L^p$</td>
</tr>
<tr>
<td>$N^v = {R^v, L^v}$</td>
<td>Virtual request, composed of set of virtual elements $R^v$ (e.g., routers, switches, VMs) and links $L^v$</td>
</tr>
<tr>
<td>$S \in \mathbb{N}$</td>
<td>Set containing all possible locations a router may have. Locations are represented by a natural number</td>
</tr>
<tr>
<td>$S^p \in R^p \times \mathbb{N}$</td>
<td>Indicates the specific location a physical device (e.g., routers, VMs) is placed on the infrastructure</td>
</tr>
<tr>
<td>$S^v \in R^v \times \mathbb{N}$</td>
<td>Indicates the location required by virtual elements belonging to virtual requests</td>
</tr>
<tr>
<td>$C_i^p \in \mathbb{N}$</td>
<td>CPU capacity of physical device $i$</td>
</tr>
<tr>
<td>$M_i^p \in \mathbb{N}$</td>
<td>Memory capacity of physical device $i$</td>
</tr>
<tr>
<td>$B_{i,j}^{k,l} \in \mathbb{N}$</td>
<td>Bandwidth capacity of physical link $(ij)$</td>
</tr>
<tr>
<td>$C_i^v \in \mathbb{N}$</td>
<td>Required CPU capacity by virtual element $i$ belonging to virtual request $r$</td>
</tr>
<tr>
<td>$M_i^v \in \mathbb{N}$</td>
<td>Required memory capacity by virtual element $i$ belonging to virtual request $r$</td>
</tr>
<tr>
<td>$B_{i,j}^{k,l} \in \mathbb{N}$</td>
<td>Required bandwidth capacity by virtual link $(ij)$ belonging to virtual request $r$</td>
</tr>
<tr>
<td>Variables</td>
<td></td>
</tr>
<tr>
<td>$A_{i,j,r,k}^l \in [0, 1]$</td>
<td>Router allocation</td>
</tr>
<tr>
<td>$A_{i,j,r,k,l}^1 \in [0, 1]$</td>
<td>Link allocation</td>
</tr>
</tbody>
</table>

The variables are outputs of our model and represent the optimal solution of the virtual network embedding problem for the given set of inputs. These variables indicate in which location the requested virtual routers and links are allocated on the physical substrate. If a request is accepted, each of its virtual routers is mapped to a physical router, whereas each virtual link is mapped to one or more consecutive physical links (a path).

- $A_{i,j,r,k}^l \in [0, 1]$ – Router allocation, indicates whether physical router $i$ is hosting virtual router $j$ from virtual network $r$.
- $A_{i,j,r,k,l}^1 \in [0, 1]$ – Link allocation, indicates whether physical link $(i, j)$ is hosting virtual link $(k, l)$ from virtual network $r$.

Based on the above inputs and outputs, we now present the objective function and its constraints. The objective function of the model aims at minimizing the total bandwidth consumed by virtual networks embedded on the substrate. The purpose of each constraint is explained next.

Objective:

$$\min \sum_{(i,j) \in E} \sum_{r \in N^p, k, l \in L^v} A_{i,j,r,k}^l B_{i,j}^{k,l}$$

Subject to:

$$\sum_{r \in N^p, k \in L^v} C_i^p A_{i,j,r,k}^l \leq C_i^p \quad \forall i \in R^p$$ (C1)

$$\sum_{r \in N^v, j \in L^v} M_i^v A_{i,j,r,k}^l \leq M_i^v \quad \forall i \in R^p$$ (C2)

$$\sum_{r \in N^v, j, l \in L^v} B_{i,j}^{k,l} A_{i,j,r,k,l}^1 \leq B_{i,j}^{k,l} \quad \forall (i, j) \in L^v$$ (C3)

$$\sum_{i \in R^p} A_{i,j,r,k}^l = 1 \quad \forall r \in N^v, j \in L^v$$ (C4)

$$\sum_{j \in L^v} A_{i,j,r,k}^l \leq 1 \quad \forall i \in R^p, r \in N^v$$ (C5)

$$\sum_{j \in L^v} A_{i,j,r,k}^l - \sum_{j \in L^v} A_{i,j,r,k,l}^1 = A_{i,r,k}^R - A_{i,r,k}^L$$ \quad (C6)

$$\forall r \in N^v, (k, l) \in L^v, i \in R^p$$

$$jA_{i,j,r,k}^R = \mathbb{I}_{A_{i,j,r,k}^R} \quad \forall (i, j) \in S^p, r \in N^v, (k, l) \in S^v$$ (C7)

$$A_{i,r,k}^R = E_{i,j}^R \quad \forall (i, j) \in E^R$$ (C8)

$$A_{i,j,r,k,l}^1 = E_{i,j,r,k,l}^1 \quad \forall (i, j, r, k, l) \in E^1$$ (C9)

Constraint set C1 ensures that the CPU capacity of each physical router will not be exceeded, therefore assuring that the CPU requirement of each virtual router will be met. Constraint set C2 applies the same restriction to the memory capacity of routers, and constraint set C3, to link bandwidth. Constraint set C4 ensures that all virtual routers will be mapped to a physical router. In turn, constraint set C5 prevents multiple virtual routers that belong to a single virtual network from sharing the same physical router. As our objective function aims at minimizing bandwidth usage, the absence of this constraint would encourage a significant number of routers from a single virtual network to share the same physical router, which could lead to availability issues. Constraint set C6 ensures that all virtual links will be mapped to a valid physical path. Thus, the physical path hosting a virtual link $(k, l)$ is guaranteed to be a valid path between the physical router hosting virtual router $k$ and the physical router hosting virtual router $l$. Internally, this equation works as follows. It ensures that an outgoing link from the physical router hosting virtual router $k$ is selected as part of the path that will host virtual link $(k, l)$. In this case, the right side of the equation equals 1 (as $A_{i,r,k}^R$ is set to 1 and $A_{i,r,k}^L$ is set to 0), forcing $A_{i,j,r,k}^l$ to be set to 0. Likewise, it ensures an incoming link to the physical router hosting virtual router $l$ is also selected. In this case, the right side of the equation equals 1, this time forcing $A_{i,j,r,k}^l$ to be set to 1. Moreover, it selects intermediate links between the aforementioned ones, thus creating a valid path from the source of the link to its destination. This happens because, in these cases, the right side of the equation equals 0 (as both $A_{i,r,k}^R$ and $A_{i,r,k}^L$ are set to 0), forcing both $A_{i,j,r,k}^l$ and $A_{i,j,r,k,l}^1$ to be set to 1. Constraint set C7 makes sure that all virtual routers with location requirements are mapped to physical routers at the required locations. This is achieved through an equality between the physical location of node $i$ (i.e., location $j$) and the required location of virtual node $k$ (i.e., $l$). Note that this equality is only enforced when variable $A_{i,j,r,k}^l$ is set to 1. Finally, constraint sets C8 and C9 ensure that all elements from previously embedded virtual networks remain hosted on the same physical elements. Router mappings are maintained by constraint set C8, while link mappings are maintained by constraint set C9. In other words, the model presets certain values in variables $A^R$ and $A^L$ in advance according to previous mappings in order to ensure they remain the same.

In spite of the differences in the embedding problem when dealing with virtual requests in ISP and datacenter networks, the above formulation is general enough to cope with both scenarios. In general, the embedding of VDC requests in datacenters does not take into account constraint set C5 (Zhang et al., 2014) and, therefore, the solution allows more than one virtual element (e.g., switches and VMs) be mapped on top of the same physical device. This relaxation is considered so as to approximate the current model with the ones observed in practice and does not compromise the obtained results.

5. Evaluation setup

In order to evaluate the impact of different network topologies
in the process of virtual network embedding, the model formalized in the previous section was implemented and run in CPLEX Optimization Studio\(^3\) version 12.3. All experiments were performed on a machine with four AMD Opteron 6276 processors and 64 GB of RAM, using the Operating System Ubuntu GNU/Linux Server 11.10 x86_64. Next, we first define the workloads used for the experiments carried out considering different InP networks. Then, we describe and discuss the results obtained for each of the network scenarios.

5.1. Workloads

To perform the experiments we adopt a strategy in line with related work, such as Yu et al. (2008), Houidi et al. (2011), and Zhang et al. (2014). Like them, we rely on time units and distribution models for the arrival and duration of requests.

We developed a virtual network request generator, which is run for a period of 500 time units for each experiment. Within each time unit, five requests are generated on average, according to a Poisson distribution. Each request has a limited duration, i.e., after a number of time units, it is removed from the substrate. Requests have an average duration of five time units, following an exponential distribution.

Physical substrate topologies for ISP networks are created using the IGen\(^4\) tool, as publicly available real substrate networks are not normalized in terms of network size. In order to generate networks with the topological features of the previously presented backbone classes – star, ladder and hub & spoke – we used, respectively, the methods Mentor, MultiTour and TwoTree. In consonance with the topology characterization presented in Section 3.1, ladder network nodes have an average degree of 3 and normalized maximum degree of 4. Star networks have a proportion of highly interconnected nodes (hubs) of less than 0.25, while in hub & spoke networks this proportion is equal or higher than 0.25. This ratio is defined as the number of nodes with connectivity degree greater than the average connectivity degree of the network, divided by the total number of network nodes. Besides these topological properties, physical networks have 50 routers, each with a total CPU capacity of 100% and 256 MB of memory. Routers are uniformly distributed among 16 locations, and the bandwidth of physical links is uniformly distributed between 1 and 10 Gbps.

In turn, the topology of each virtual network requested to ISP providers is generated using BRITE\(^5\) with the Barabási–Albert (BA-2) model (Albert and Barabási, 2000). Each virtual network has between 2 and 5 routers. Virtual routers require between 10% and 50% of CPU and between 24 MB and 128 MB of memory. Both parameters follow a uniform distribution. Virtual link bandwidth is uniformly distributed between 1 and 5 Gbps.

With respect to datacenter network topologies, they are created following the topologies described in Section 3.2, namely, conventional multi-rooted tree (Bari et al., 2013), VL2 (Greenberg et al., 2009), and Fat-Tree (Leiserson, 1985). Conventional tree topologies are organized with 2 core routers, 10 aggregation switches, and 30 access switches. Each core router is connected to 5 aggregation switches which, in turn, are connected to 15 access switches each. VL2 topologies are designed with three layers (or stages) of 14 switches. Each switch in a layer is connected to all switches in the previous and next layers. Last, Fat-tree topologies are designed with 6 pods of switches (i.e., k = 6). Consequently, Fat-tree topologies have 6 core switches, which are connected to every pod. All physical datacenter network topologies have exactly 42 routers/switches. Additionally, each topology has 120 physical machines distributed in the access layer (which is in line with the work by Zhang et al., 2014). Each physical machine has a total CPU capacity of 400% (4 CPU cores), 8 GB of memory, and contains a 5 Gbps network adapter. Physical link bandwidth in the core layer is set to 10 Gbps, while in the aggregate layer it is set to 5 Gbps. Switches in the access layer are uniformly distributed amongst 16 locations.

Datacenter requests are topologically different from the ones submitted to ISPs, as they usually represent topology of distributed applications (e.g., MapReduce). We adopt the same strategy of Zhang et al. (2014) to generate them. VDC request topologies follow the conventional multi-rooted tree design. The number of core nodes is fixed to 1, the aggregation layer has between 1 and 3 nodes, and the access layer has between 1 and 5 nodes. Each VDC node represents a virtual machine, which requires between 40% and 200% of CPU and between 819 MB and 4096 MB of memory (i.e., demanded virtual resources are between 10% and 50% of the physical resources capacity). Virtual link bandwidth is uniformly distributed between 500 and 1000 Mbps. We also ensure that links on upper layers of VDC requests have more bandwidth than those in lower levels. Hence, it is possible to generate a wide range of VDC requests with different requirement in terms of topologies and resources.

Two scenarios were evaluated on each physical topology. The distinctive feature of each scenario is the presence or absence of location requirements. In the first scenario, each virtual request has two endpoints (routers in VNs and virtual machines in VDCs) with location requirements which are randomly selected amongst the 16 locations. In the second scenario, there are no such requirements. Each experiment was run 30 times, considering different instances for each type of network substrate. All results have a confidence level of 90% or higher.

6. Results

First, we analyze the rejection rate of virtual requests (i.e., VNs and VDCs) for the scenario described in Section 5. As previously mentioned in the Introduction, the rejection rate is defined as the ratio between the number of rejected requests and the number of incoming ones. Virtual requests are only rejected if it is not possible to map all of their routers and links on the physical substrate. Likewise, virtual datacenter (VDC) requests are only rejected if it is not possible to map all of their switches, VMs, and links on the infrastructure.

6.1. Rejection rates

Location requirements and network connectivity play a significant role in the rejection rate. Figs. 4 and 5 depict the average rejection rate of virtual requests considering, respectively, ISP and datacenter networks. Each point on the graphs represents the average rejection rate since the beginning of the experiment until the current time unit. It is clear that when location requirements are considered, rejection rates are substantially high (ranging from 65.38% to 83.71% in ISP networks and from 35.91% to 56.31% in datacenter networks) for all three physical topologies. When location requirements are relaxed, rejection rates are comparatively lower, ranging from 0% to 41.32% in ISP networks and from 32.54% to 45.83% in datacenter networks. The observed variation in rejection rates (in each scenario) is mainly due to the reduction in the exploration space of feasible solutions caused by the presence of location constraints.

The graphs depicted in Figs. 4 and 5 also reveal that there is considerable difference in rejection rates when using different

\(^3\) http://www-01.ibm.com/software/integration/optimization/cplex-optimization-studio/
\(^4\) http://igen.sourceforge.net/
\(^5\) http://www.cs.bu.edu/brite/
physical topologies belonging to the same topological class (i.e., ISP or datacenter). Considering the ISP topology class, hub & spoke networks lead to a lower rejection rate in comparison to other topologies in both evaluated scenarios (68.44% in the scenario with location requirements and 0.53% in the scenario without such requirements). In contrast, star topology networks lead to the worst performance (rejection rate of 85.04% in the scenario with location requirements and 43.10% in the scenario without location requirements). Ladder topology networks present rejection rates of 75.63% and 24.66% for the scenarios with and without location requirements, respectively. Hub & spoke networks tend to cause the rejection of a lower number of requests because they have, on average, a higher number of highly interconnected nodes (hubs). The presence of multiple hubs lowers the probability that the depletion of the resources of one of these central nodes may cause a significant impact on the ability to embed future requests. In contrast, as star topology networks have a low number of central nodes, there is a high probability that these nodes may become a bottleneck in the process of virtual network embedding if their resources are depleted. In ladder networks, as there are no central nodes, the depletion of resources in some physical links may hinder the creation of virtual links that would use such physical links as “bridges” to interconnect certain points of the infrastructure.

Similarly, there is also a perceptible difference in rejection rates when considering distinct datacenter topologies. Fat-Tree topology networks lead to the worst performance in relation to other topologies for both evaluated scenarios (rejection rate of 56.31% in the scenario with location requirements and 45.83% in the scenario without location requirements). In turn, VL2 networks lead to rejection rates of 35.91% and 32.54% for the scenarios with and without location requirements, whereas conventional networks, to rejection rates of 49.13% and 38.06%, respectively. VL2 topologies tend to cause the rejection of a lower number of requests because each layer is fully connected to each other, i.e., each switch in a layer is connected to all switches belonging to adjacent layers. Thus, bandwidth depletion of some physical links bridging different layers does not hinder the embedding process since there is a significant number of unused paths (or with enough residual bandwidth) that interconnect sets of available resources. In contrast, as Fat-Tree topologies are not fully connected as VL2, the number of links amongst layers (particularly between access and aggregation layers) tends to be smaller in comparison to other topologies. Thus, there is a high probability that these links become a bottleneck in the embedding process if such resources are depleted. Conventional topologies represent an intermediate case between VL2 and Fat-Tree in terms of connectivity, i.e., they are neither too wired as VL2 nor have its resources too scattered as Fat-Tree. Another observation worth highlighting is that rejection rates do not vary considerably in datacenter network topologies when evaluating the scenario without location requirements (~13%). We emphasize that this behavior is primarily influenced by the inherited characteristics of the VDC model, which allows VMs of the same request to be mapped on top of the same physical machine (i.e., the embedding model does not consider constraint C5). Besides, topological properties of datacenter networks also contribute to this behavior, as we will discuss later.
6.2. Bandwidth overhead

Bandwidth overhead is significantly lower on network topologies that have multiple hubs. Figs. 6 and 7 illustrate the average overhead caused by virtual requests embedded in each experiment. This overhead is measured as the ratio between the effective bandwidth consumed by virtual requests hosted on the physical substrate and the bandwidth requested by them. In general, the actual bandwidth consumption is higher than the total bandwidth required by virtual requests, due to the frequent need to map virtual links to paths composed of multiple physical links. The absence of overhead is observed only when each virtual link is mapped to a single physical link (ratio of 1.0), or when endpoints of a virtual link (e.g., virtual machines) are mapped on top of the same physical machine in the datacenter infrastructure (ratio <1.0 – in this case, virtual links do not consume effective bandwidth). We emphasize that lower overhead rates directly favor the infrastructure provider, by sparing resources that may be used to embed future requests. Moreover, this lower resource consumption may lead to lower costs for requesters. With respect to ISPs, ladder topologies lead to higher average overheads in comparison to other topologies (94.59% in scenarios with location requirements and 23.36% in scenarios without location requirements). This is due to the fact that this topological structure has, on average, longer distances (in terms of hops) between nodes, in addition to the absence of hubs. Hub & spoke networks achieve the lowest overhead rates (64.67% in the scenario with location requirements), as they have a higher number of hubs and interconnections between nodes. In the scenario without location requirements (Fig. 6(b)), the average overhead rates observed in the ladder and star topologies are similar – 23.36% and 21.92%, respectively – whereas the overhead rate caused by the employment of the hub & spoke topology is only 1.89%.

With respect to datacenter networks, Fat-Tree topologies lead to a higher bandwidth overhead in comparison to other topologies in both scenarios (133.62% in scenarios with location requirements and 91.81% in scenarios without location requirements). In contrast, VL2 networks lead to a lower bandwidth overhead in both evaluated scenarios (66.48% in the scenario with location requirements and 63.26% in the scenario without such requirements). Conventional topologies exhibit an intermediate overhead between VL2 and Fat-Tree (112.47% in the scenario with location requirements and 84.28% in the scenario without such requirements). These comparatively higher overheads observed in datacenter network topologies is mainly due to two factors. First, the depletion of bandwidth on links connecting access and aggregation layers contributes to embedding solutions consisting of longer paths (which might include links of the network core). As a consequence, physical topologies with a fewer number of links between access and aggregation layers tend to be further impacted by exhaustion. Second, as VMs are assigned to physical machines attached to access switches, any virtual link between VMs hosted on different physical machines uses at least two physical links. Obtained results demonstrate that, when location requirements are considered, virtual links tend to be mapped to longer paths on the substrate, due to a higher average distance between the locations where virtual elements are expected to be hosted. However, it is noteworthy that the usage of both ISP topologies (that

![Fig. 6. Average bandwidth overhead needed to accommodate accepted VN requests.](image)

![Fig. 7. Average bandwidth overhead needed to accommodate accepted VDC requests.](image)
have a higher number of hubs) and datacenter topologies (that have a higher number of links – especially between access and aggregation layers) tends to reduce the impact of these location requirements on the bandwidth overhead.

6.3. CPU and memory consumption

The rejection of virtual requests is not in general caused by the global exhaustion of physical resources (CPU and memory). We evaluate the average consumption of physical resources – CPU and memory capacity of physical devices (routers and physical machines), as well as link bandwidth – in all experiments. We emphasize that higher overall resource consumption may lead to lower costs for requesters. This is due to the fact that, as providers are able to better utilize available resources, costs (i.e., CAPEX and OPEX) are amortized over a potentially higher number of customers. Figs. 8 and 9 show the average CPU consumption of physical devices in InP networks. In ISP networks, considering location requirements, we observe that the average CPU consumption when employing the ladder topology (11.49%) is approximately twice as high as the average CPU consumption in the experiment in which we use the star topology (5.36%). The hub & spoke topology led to an average CPU consumption of 16.82%. For the scenario without location requirements, the average CPU consumption in the star network is 23.80%, whereas in the ladder and hub & spoke networks, 35.32% and 38.60%. In datacenter networks, VL2 topologies lead to a higher average CPU consumption in comparison to other datacenter topologies in both evaluated scenarios (51.05% in the scenario with location requirements and 65.63% in the scenario without such requirements). In contrast, Fat-Tree topologies lead to a lower average CPU consumption in both evaluated scenarios (38.40% in the scenario with location requirements and 53.82% in the scenario without such requirements). Conventional topology networks present an average CPU consumption of 43.40% and 54.56% for the scenarios with and without location requirements, respectively. Increased consumption of CPU resources on both network topology classes (ISP and datacenter) is directly influenced by the number of virtual elements embedded on the substrate. This explains the higher consumption in scenarios that do not consider location requirements, as rejection rates are lower in these cases. Further, we emphasize that the comparatively higher CPU consumption observed in all datacenter network scenarios is also due to the greater flexibility in embedding VDC requests (by virtue of the absence of constraint C5). However, the relatively low CPU consumption, which does not exceed 40% in ISP networks and 65% in datacenter networks, reveals that the rejection of virtual requests is not caused by the global exhaustion of CPU resources in the infrastructure.

In Figs. 10 and 11, we present the average memory consumption of physical devices. The behavior of memory usage in such devices is similar to that of CPU usage. Considering location requirements in ISP networks, the average memory usage is of 5.65% in the star topology, 11.21% in the ladder topology, and 15.93% in the hub & spoke topology. In contrast, experiments that do not consider such requirements lead to an average memory utilization of 23.49% in the star topology, 35.52% in the ladder topology, and 38.03% in the hub & spoke topology. In datacenter networks, considering location requirements, the average memory usage is of 38.06% in the Fat-Tree topology, 41.97% in the conventional topology, and 51.14% in the VL2 topology. Experiments that do not consider such requirements lead to an average memory utilization of 53.33% in the Fat-Tree topology, 54.41% in the conventional topology, and 65.75% in the VL2 topology. Note that memory resources are not entirely depleted in any of the experiments. Thus, we can state that this factor, similarly to CPU usage, is not the cause of the rejection rates observed in both InP networks.

6.4. Bandwidth consumption

The overuse of bandwidth in specific physical links hinders the embedding of a higher number of virtual requests. The average bandwidth consumption in physical links is shown in Figs. 12 and 13. Considering the ISP network experiments in which virtual networks have location requirements, average link utilization is of 18.01% in the star topology, 40.63% in the ladder topology, and 35.89% in the hub & spoke topology. In the experiments without location requirements, the average consumption is 44.93% in the star topology, 66.50% in the ladder topology, and 54.08% in the hub & spoke topology. With respect to datacenter networks in the scenarios with location requirements, the average link utilization is 19.18% in VL2 topology, 43.37% in conventional topology, and 52.39% in Fat-Tree topology. In the experiments without location requirements, the average consumption is 20.97% in VL2 topology, 54.44% in Fat-Tree topology, and 44.02% in conventional topology. These results show that the overall bandwidth consumption in the physical network cannot also be responsible for the rejection rates observed in the experiments. The saturation of specific points in the physical infrastructure makes it impossible to embed a higher number of virtual requests, even though considerable amount of spare resources are, from a global point-of-view, available.

ISP and datacenter network topologies may have different connectivity degrees and, due to the depletion of resources in specific parts of the substrate (e.g., links connected to hubs or bridges), infrastructure partitioning occurs at different levels. In this context, a partition is defined as a strongly connected component (SCC) with residual bandwidth equal to or greater than the
average link bandwidth requested by virtual requests. In ISP networks, we observed that, in scenarios with location requirements, **star** topologies have, upon the embedding of a new VN, 13.60 partitions with an average of 3.63 routers per component, while **ladder** topologies have, on average, 12.30 partitions with 4.38 routers per component. In turn, **hub & spoke** topologies present lower levels of partitioning and a higher number of routers per component, i.e., 7.22 partitions and 7.52 routers, respectively. In scenarios without location requirements, infrastructure partitioning levels are higher due to higher link bandwidth consumption. **Star** and **ladder** topologies have similar average partitioning levels; 19.20 and 19.67 partitions with 2.61 and 2.48 routers, respectively. **Hub & spoke** topologies have, on average, 12.65 partitions and 3.94 routers per component. Datacenter network topologies present, in general, larger partitions in both scenarios for the three physical topologies considered. In scenarios with location requirements, **VL2** and **conventional** topologies have similar average partitioning levels, i.e., 20.99 and 19.49 partitions with 7.74 and 8.38 devices (including routers and physical machines), respectively. In turn, **Fat-Tree** topologies have, on average, 16.89 partitions with 9.66 devices each. In scenarios without location requirements, **VL2** topologies have, on average, 3.149 partitions with 5.16 devices. **Conventional** and **Fat-Tree** topologies have, on average, 22.83 and 26.29 partitions with 7.14 and 6.15 routers, respectively. Next, we analyze how this and other factors influence the rejection rates observed.

### 6.5. Analyzing rejection causes

There is a direct correlation between partitioning levels and rejection rates in ISP networks. The same is not valid in datacenter networks as rejection rates usually result from a set of combined factors. Figs. 14 and 15 show results regarding causes of rejection in the performed experiments. Three main causes of rejection have been observed: the absence of a valid partition, the lack of physical router (or physical machine) resources within a partition, and other topological factors. The first cause is related to the unavailability of a proper partition on the infrastructure with size equal to or greater than the number of virtual elements requested. This means that it is not possible to map the virtual requests (VN or VDC) on the topology due to the lack of connectivity between physical devices. In other words, there is no SCC that has, for each pair \((a,b)\) of physical devices (routers or virtual machines), a path between “a” and “b” with bandwidth greater than or equal to the maximum bandwidth requested. In ISP networks, considering location requirements, 59.45% of rejections on **star** topologies as well as 54.44% on **ladder** topologies are caused by this kind of connectivity problem. In scenarios without location requirements, 99.62% of rejections on **star** topologies and 99.47% of rejections on **ladder** topologies are associated with this cause. In **hub & spoke** topologies, rejections associated with this cause amount to 20.85% in scenarios with location requirements and 28.50% in scenarios without such requirements. In turn, datacenter network topologies present a negligible percentage of rejections associated with this cause. **Fat-Tree** is the only datacenter network topology that has rejections associated with this cause (14.89% in the scenario with location requirements and 4.69% in the scenario without such requirements). These results suggest a direct correlation between partitioning levels and rejection rates in ISP networks. Instead, this correlation is not precisely valid to datacenter networks as, on
average, these topologies are more connected and tend not to be as partitioned as ISP networks. In this case, the rejection cause is mainly determined by other factors as we show next.

The second cause of rejection—the lack of physical device resources within a partition—occurs when there is an appropriate partition (i.e., there is an SCC that contains enough physical devices to map all virtual elements, but it is not possible to map the virtual request due to the depletion of physical resources (CPU and memory) in this partition. Considering ISP network topologies in the scenario with location requirements, rejection rates caused by insufficient resources are 3.2%, 1.8%, and 14.7% on star, ladder, and hub & spoke topologies, respectively. In scenarios without such requirements, these rates are, respectively, 0.33%, 0.37%, and 2.58%. Star and ladder topologies present similar percentages in both scenarios. In scenarios without location requirements, an insignificant number of VNs are rejected due to this cause, as the main cause of rejection is infrastructure partitioning. In hub & spoke topologies, when considering location requirements, the percentage of VN rejections associated with this cause is higher in contrast to other topologies. This behavior is related to lower rejection rates associated with the absence of valid partitions in this topology, which leads to greater resource usage in physical elements.

With respect to datacenter network topologies in the scenario with location requirements, we observe that rejection rates caused by insufficient physical resources are 47.92%, 92.60%, and 98.59% on Fat-tree, conventional, and VL2 topologies, respectively. In scenarios without such requirements, these rates are, respectively, 6.07%, 27.06%, and 3.35%. We can observe that the percentage of rejections associated with this cause is notably higher in scenarios with location requirements. In this scenario, virtual requests use physical resources (CPU and memory) from different partitions, which directly impact the overall fragmentation level of physical resources in the infrastructure. On the other hand, when location requirements are not considered, this cause tends to not impact as deeply as in the scenario with such requirements.

The third cause of rejection refers to virtual requests that are denied despite the existence of an adequate partition and available resources on its physical devices. In this case, rejections are a result of topological factors within a partition that make it impossible to map an incoming virtual request. This occurs either when location constraints cannot be met or when virtual requests have a significantly different topology in relation to the topology available in the partition. In the latter situation, a higher amount of resources will be required to embed these requests, which may not be available. In ISP networks, when considering location requirements, star and ladder topologies have, respectively, 37.30% and 43.71% of rejection rates associated with topological properties. In scenarios without location requirements, these percentages are 0.04% and 0.15%, respectively. Rejections related to topological factors in hub & spoke networks amount to 64.39% in scenarios with location requirements, and 68.90% in scenarios with no such requirements. When analyzing rejections associated with topological factors, star and ladder topologies present a similar behavior, as partitioning levels and overall rejection rates are similar in both topologies. Nevertheless, minimal differences exist, which can be explained by topological factors such as the absence of hub nodes in ladder networks. In hub & spoke networks, however, topological factors are the most significant cause of rejection rates. Although this type of topology presents higher connectivity, the rejection of a lower number of VNs leads to higher bandwidth usage in

![Fig. 11. Average memory usage of physical machines in datacenter networks.](image)

![Fig. 12. Average bandwidth usage of physical links on ISP networks.](image)
physical links, which leads to reduced connectivity within network partitions. In datacenter network topologies, considering location requirements, we observe that rejection rates resulting from this cause are 1.40%, 7.39% and 37.18% on VL2, conventional and Fat-Tree topologies, respectively. In scenarios without such requirements, these rates are, respectively, 72.93%, 96.64% and 89.22%. When analyzing rejections associated with topological factors in datacenter networks, we observe that location requirements do not impact rejection rates as much as in ISP networks. Most of this behavior is influenced by the high connectivity of all datacenter network topologies. However, when location requirements are not considered, the percentage of rejection associated with this cause is substantially higher. As this scenario leads to a lower rejection of virtual requests, there is a higher bandwidth usage in physical links, which reduces the connectivity between adjacent layers (especially between access and aggregation layers).

### 6.6. Time consumption

The time needed to find an optimal mapping tends to be higher when mapping virtual requests on highly interconnected network topologies. In Figs. 16 and 17 we present the average time needed to find the optimal mapping of each accepted request. We emphasize that, in this graph, the vertical axis is shown in logarithmic scale, as results differ significantly. In all ISP scenario with location requirements, the time needed to optimally embed virtual requests remains under 3 s. The star topology leads to an average solution time of 0.13 s, the ladder topology to an average time of 0.92 s, and the hub & spoke topology, to an average time of 2.41 s. In turn, the time to find the optimal mapping in datacenter network topologies is substantially higher. Considering location requirements, the use of Fat-Tree topology leads to an average of 31.95 s, the conventional topology to an average of 55.27 s, and the VL2 topology to an average of 88.40 s. The time consumption is, on average, higher when location requirements are not considered. The Fat-Tree topology leads to an average of 37.17 s, the conventional topology to 59.95 s, and the VL2 topology to 283.61 s.

The relatively high solution times observed in scenarios that employ the hub & spoke and VL2 topologies are explained by the presence of a higher number of links in this topology. Likewise, the removal of location requirements also leads to higher solution times as in this case the space of feasible solutions gets larger. Furthermore, a number of peaks are observed near the beginning of the experiments, reaching a maximum of 24.8 s in the scenario that employs the ladder topology and 542.43 s in the scenario in which the VL2 topology is employed. This behavior can be explained by the larger amount of unused resources in the beginning of these experiments (when the availability of resources on physical networks is substantially high), which also increases the number of possible mappings for virtual networks on the substrates.

### Fig. 13. Average bandwidth usage of physical links on datacenter networks.

### Fig. 14. Average percentage of rejection causes observed in all ISP experiments.
6.7. Summary

The results presented in this section show that the employment of different types of physical network topologies in network virtualization environments causes a significant impact on rejection rates and physical resource usage. This impact is even more pronounced when location requirements of virtual requests are considered. Such experiments reveal that the rejection of virtual requests is not caused by the overall depletion of resources in the infrastructure. Instead, it is caused by factors related to certain topological features.

The main factor that influences the rejection of virtual requests is resource depletion in specific regions of the substrate, which leads to higher partitioning levels in the infrastructure. For example, in star and hub & spoke ISP topologies, the exhaustion of resources in physical hubs, as well as links connected to them, tends to be the main cause that makes it impossible to map new virtual network requests even when there are partitions with sufficient size and connectivity. In ladder topologies, in which there are no hubs, the main cause for the increase in rejected requests is the depletion of resources in specific connections between nodes. Links used as “bridges” to interconnect different
points of the infrastructure can become bottlenecks. If the bandwidth of one of these links is sufficiently depleted, the infrastructure is partitioned into two groups of routers with no connectivity between them. In contrast, the observed rejections of virtual requests in datacenter networks are primarily associated with the lack of physical resources in a partition (i.e., there is no available resources at all, or they are too fragmented) or with the inherent inability to accommodate certain requests due to topological incompatibilities. Moreover, we emphasize that in all studied topologies (both in ISP or datacenter networks), partitioning may lead to the existence of segments that, despite having adequate size and sufficient resources, are still unable to accommodate certain virtual requests due to topological factors.

In summary, mapping virtual requests in star or Fat-Tree topologies leads to low solution times but also high rejection rates, and consequently, to low physical resource usage. In ladder or conventional topologies, rejection rates and solution times have intermediate values in relation to other topology classes, but embedded virtual requests tend to consume a greater amount of bandwidth. Finally, hub & spoke and VL2 topologies lead to low rejection rates and average bandwidth overhead, but the time needed to find the optimal solution is comparatively higher.

7. Conclusion

Network virtualization is a topic that has received considerable attention from both the scientific community and the Industry, resulting in a series of studies involving, predominantly, issues related to virtual network embedding. However, as far as we are aware of, there have been no previous attempts to evaluate the result of mapping strategies considering different network topologies in a precise manner. Previous work, in general, use organic or generic topologies, which do not faithfully represent the topological properties present in real InP networks.

After formalizing an optimal online virtual network embedding model and applying it on substrates with different topological features that are typically present in ISP and datacenter networks, we characterized the impact of different types of topologies regarding rejection rates and physical resource usage. The obtained results evidence the significant impact caused by embedding virtual requests on physical substrates with different topological features. The ability to embed virtual requests is hindered by resource depletion in some specific points of the physical infrastructure, although a global view of the network reveals that there are still resources available in the remainder of the substrate. This impact is even more considerable when the embedding model considers location requirements of virtual requests. In short, the results reveal that there is a relationship between rejection rates and resource consumption. Network topologies that are intrinsically more connected tend to reject a lower number of virtual requests (e.g., hub & spoke and VL2 topologies) and, consequently, tend to incur comparatively higher resource utilization. From this study, we provide consistent insights on how a physical network topology affects virtual network embedding quality. As a future work, we envision to assess the relationship between the involved costs (i.e., CAPEX and OPEX) and the revenue obtained by leasing resources through network virtualization on each network topology class.

Acknowledgment

This work has been partially supported by FP7/CNPq (Project SecFuNet, FP7-ICT-2011-EU-Brazil), RNP-CTIC (Project ReVir), as well as PRONEM/FAPERGS/CNPq (Project NPRV).

References

Yu, M., Yi, Y., Redford, J., Chiang, M., 2008. Rethinking virtual network embedding:

