## Copyright Notice



Institute of Communication Networks and Computer Engineering
University of Stuttgart
Pfaffenwaldring 47, D-70569 Stuttgart, Germany
Phone: ++49-711-685-68026, Fax: ++49-711-685-67983
Email: mail@ikr.uni-stuttgart.de, http://www.ikr.uni-stuttgart.de

# Modeling Dynamic Traffic Demand Behavior in Telecommunication Networks

Tobias Enderle, Uwe Bauknecht
Institut für Kommunikationsnetze und Rechnersysteme (IKR), Universität Stuttgart, Stuttgart, Germany,
{tobias.enderle, uwe.bauknecht}@ikr.uni-stuttgart.de

## Kurzfassung

Technologien wie 5G und IoT werden neue, volatilere Netzdienste ermöglichen, deren Verhalten das Verkehrsaufkommen in künftigen Kommunikationsnetzen erheblich verändert, sodass diese zunehmend flexibler und effizienter betrieben werden müssen. Um die Auswirkungen dieser Entwicklungen untersuchen zu können, werden ausreichend präzise Modellierungsansätze benötigt, welche das dynamische Verhalten unterschiedlicher Netzdienste abbilden können. In dieser Arbeit wird ein derartiger Ansatz vorgestellt und dessen Anwendbarkeit zur Untersuchung von Weitverkehrsnetzen anhand einer Beispielstudie demonstriert. Gegenstand dieser Studie ist die Ermittlung des Ressourcenbedarfs eines Multi-Layer-Netzes durch Optimierungsheuristiken.

## Abstract

Technologies such as 5G and IoT will give rise to novel network services that will drastically change the behavior of traffic demands in future telecommunication networks which will have to become more flexible and more efficient to deal with the increasing traffic volatility. In order to study the effects of these changes, sufficiently precise modeling approaches which reflect the dynamic behavior of diverse network services are required. In this work we present such an approach and demonstrate its applicability to the study of wide area networks by an example evaluating the resource requirements of a multi-layer network through optimization heuristics.

## 1 Introduction

Internet service providers (ISPs) today face a number of social, economic and technological factors which are expected to drastically change the network landscape of the near future. In the private sector the number of connected devices is growing rapidly through the widespread adoption of streaming appliances, smart home applications and wearables while big data analytics, cloud computing and cloud storage are becoming essential tools in the business sector. Furthermore, novel use cases such as the Internet of Things (IoT) and augmented or virtual reality applications are expected to be among the fastest growing traffic contributors while requiring low end-to-end delay. In terms of volume, however, inter-data center traffic will be the dominant component of future traffic accounting for more than 50 % of all IP traffic in 2021. The second largest component will be video services which are forecast to account for about 39 % of all IP traffic, and will continue to grow at a rate of 26 % per year[1] [1]. In combination with new access-technologies such as FTTx and 5G, which will increase data rates and promise significantly reduced delays [2], the traffic in ISP transport networks will not just increase in volume, but also become much more volatile and delay-sensitive.

For the development of network architectures and algorithms, that are capable of handling future network traffic requirements, a suitable traffic demand model is necessary. This model must be able to deal with various service types each having different requirements in terms of Quality of Service (QoS) parameters, required data rate and possible endpoints. Due to the increased volatility of future traffic demands the model must also represent the temporal traffic behavior.

We propose a traffic demand model and algorithm for the generation of dynamic traffic demand matrices for wide area networks. The model is based on the fact that different types of services exist in a network, each having different requirements in terms of QoS and possible endpoints. For example, traffic that originates from data center replication will only exist between nodes that are connected to a data center whereas voice traffic is likely to exist between any node pair. The model allows for the flexible adjustment of service shares and overall peak traffic. This is an important requirement because it allows for the generation of demand matrices that resemble different load situations.

This paper is structured as follows. Section 2 gives a brief overview on aspects and requirements of traffic demand generators and outlines existing approaches. Section 3 describes our traffic modeling approach. Section 4 presents an example how the model can be parametrized based on publicly available data. Section 5 shows, how the model can be used in network-level simulations. Section 6 provides concluding remarks.

[1] Compound Annual Growth Rate 2016–2021

## 2 Traffic Modeling for Wide Area Networks

Obtaining traffic demand matrices is a known problem in the network community. There are sources for measured demand matrices and also various models that can be used for the synthesis of demand matrices. Publicly available demand matrices like in [3] and [4] serve as a starting point for the analysis of traffic characteristics. However, they are taken from research networks and therefore not all of their characteristics can be transferred to the traffic in a commercial ISP network, that is composed mainly of consumer and enterprise data. Also, the available traffic demand matrices do not distinguish between different service types, but only measure the overall traffic amount between the network nodes. Hence, traffic characteristics can not be derived from these measurements alone and therefore other models must be incorporated.

Traffic characteristics can be divided into two parts, namely temporal and spatial aspects [5]. A simple yet powerful approach for modeling the spatial distribution of traffic in the network is the gravity model [6]. It assumes that the amount of traffic between two nodes in the network scales proportionally with the product of the number of users connected to those nodes and reciprocally with the squared distance between them. The authors in [7] also employ the gravity model. However, they distinguish between three traffic types, namely voice, transaction data and Internet traffic, and model each of them in a slightly modified way. Both models assume symmetric traffic demands between the nodes, i. e. the volume from node $u$ to node $v$ is the same as from $v$ to $u$. On the contrary, the Independent-Connection (IC) Model in [8] models traffic as bidirectional connections with an adjustable ratio of up- and downstream traffic. This aspect is of particular interest to us because the majority of today's traffic is asymmetric in nature [9]. Take for example video streaming with a high downstream and a low upstream data rate.

Considering temporal aspects, the IC Model also introduces the idea of time-varying activity levels for each node. We will transfer this idea to the individual service types. Furthermore, we adopt the approach introduced in [10] that models the temporal traffic behavior as a dynamic cyclostationary process. This process consists of a deterministic component that models diurnal patterns and a random component that represents fluctuations over time. The diurnal behavior of Internet traffic becomes evident in the traffic traces of Internet exchange (IX) points like DE-CIX [11]. In our model, the variance of the demand's random component is determined by the number of users related to this demand. In this way, demands with a high aggregation level are less volatile than demands with a small aggregation level.

A software implementation of a generator for dynamic demand matrices has been developed in the framework of the DISCUS project [12, 13]. The generator distinguishes different services. However, demands are modeled in a bottom-up approach and the adjustment of overall traffic shares per service and peak hour traffic is not possible.

All the literature presented here provides important aspects fulfilling some of our requirements for a dynamic, diverse service-aware traffic demand model. However, to the best of our knowledge, there does not exist a single model that combines all the aspects we mentioned here.

## 3 Demand Modeling and Generation Approach

The model is intended to describe traffic demands between nodes in wide area networks. In our case, a single traffic demand denotes the maximum data rate that needs to be provided between a source node and a target node during a specific time interval. All traffic demands between any pair of nodes in a network form a traffic demand matrix. Eventually, the model is used to generate traffic demand matrices for consecutive, fixed-length intervals.

To be more precise, let the network consist of a set of nodes $V$ and further let $E = V \times V$ be the set of all node pairs. Each node $v \in V$ is annotated with its geographical location $l_v$ and corresponding time offset $\Delta_v$ w. r. t. a prespecified reference timezone. Furthermore, the number of connected users $y_v$ and a set of features $B_v$ are assigned to each node. Possible examples for features are a data center or an Internet exchange point which are connected to that node. Let $B$ be the set of all features that exist in the network, i. e. $B = \cup_{v \in V} B_v$. In general, the set of nodes that provide a feature $b \in B$ is a proper subset of $V$, i. e. not all nodes provide feature $b$. Especially in networks that consist of nodes of different aggregation levels, this is the case. Therefore, in addition to the number of directly connected users, nodes also have aggregated user counts $y_{v,b}$ for each feature $b \in B_v$. These user counts are computed with a nearest neighbor approach from the node locations $l_v$ and user counts $y_v$ of the different nodes. For a feature $b \in B$, the users $y_v$ of each node $v$ are assigned to the geographically closest node providing $b$. A node providing $b$ assigns its users to itself. Hence, the aggregated users are given by

$$y_{v,b} = \sum_{u \in V_{v,b}} y_u \tag{1}$$

where

$$V_{v,b} = \left\{ u \in V : \underset{w \in V : b \in B_w}{\arg\min}\ d((u,w)) = v \right\} \tag{2}$$

is the set of nodes in $V$ assigned to $v$ for the feature $b$ and $d((u,w))$ is the geographical distance between $u$ and $w$.

Let $S$ be the set of service types in the network, for example Video on Demand (VoD), gaming, data backup etc. Each service type is characterized by a usage profile $p_s$, a downstream traffic portion $f_s$ and its total traffic share $\rho_s$. Additionally, a service can require one feature $b_{s,\text{source}}$ at the source node and one feature $b_{s,\text{target}}$ at the target node. A traffic demand exists between a node pair only if both source and target provide the requested features. For example, VoD is typically provided through Content Delivery

Networks (CDNs). Therefore, traffic demands of a VoD service can be modeled to only exist between a node pair $(u,v)$ for which $v$ comprises a CDN data center.

Directed traffic demands are created between every two distinct nodes $u,v \in V, u \neq v$. For each discrete point in time $t \in T \subset \mathbb{N}$ the output of the model is a set of traffic demand matrices $\{X_s(t) \in \mathbb{R}_{\geq 0}^{|V|\times|V|} : s \in S\}$. Each matrix $X_s(t)$ consists of the traffic demands $x_{e,s}(t)$ of service type $s$ between the different node pairs $e = (u,v)$. As mentioned above, for all matrices the elements $x_{e,s}(t)$ on the main diagonal, i. e. with $e = (u,u)$, are zero.

The downstream traffic portion $f_s$ is a value between zero and one as defined in [8]. For a demand from node $u$ to node $v$, it defines how much traffic flows in downstream direction, i. e. from $v$ to $u$. The remaining $1 - f_s$ traffic units flow in upstream direction from $u$ to $v$. This allows us to model asymmetric services like data backups or VoD properly. The temporal behavior of the service is derived from its usage profile. This profile is a deterministic function $p_s : T \to [0,1]$ that assigns the usage of the service to each point in time $t \in T$. The application of a deterministic profile is motivated by the fact that, in wide area networks, we are dealing with highly aggregated traffic demands of hundreds or thousands of users per node. Hence, the usage profile corresponds to the average user activity. Due to the diurnal behavior of the traffic in wide area networks it is, in practice, sufficient for a usage profile to describe the time frame of one day. Since real traffic demands are not deterministic we add a random term $n_{e,s}(t)$, drawn from a normal distribution that is centered at zero, to the profile values. This approach is comparable with the dynamic cyclo-stationary model in [10]. The standard deviation $\sigma_{e,s}$ of the distribution increases linearly when the product of the number of connected users at source $u$ and target node $v$ decreases. This results in volatile traffic demands for node pairs with few users and rather smooth demands for node pairs with many users. More concretely, the standard deviation is calculated by

$$\sigma_{(u,v),s} = \frac{\sigma_{\max} - \sigma_{\min}}{\hat{y}_{\max} - \hat{y}_{\min}} (\hat{y}_{\max} - \hat{y}_{(u,v),s}) + \sigma_{\min} \tag{3}$$

where

$$\hat{y}_{\max} = \max_{u',v' \in V, b_1, b_2 \in B} y_{u',b_1} y_{v',b_2} \tag{4}$$

and

$$\hat{y}_{\min} = \min_{u',v' \in V} y_{u'} y_{v'} \tag{5}$$

are the maximum and minimum user product in the network, respectively,

$$\hat{y}_{(u,v),s} = y_{u,b_{s,\text{source}}} y_{v,b_{s,\text{target}}} \tag{6}$$

is the user product of nodes $u$ and $v$ for the requested features, $B$ is the set of all existing features and $\sigma_{\max}$ and $\sigma_{\min}$ are model parameters. The variables $y_{u,b_{s,\text{source}}}$ and $y_{v,b_{s,\text{target}}}$ represent the number of users at source and target node, respectively, aggregated according to the required node features. In case no source or target feature is specified for a service, $y_{u,b}$ and $y_{v,b}$ fall back to $y_u$ and $y_v$, respectively.

At time step $t$ the preliminary and undirected traffic demand of service $s$ between the node pair $e = (u,v)$ is calculated by

$$x'_{e,s}(t) = \max\left(\frac{(p_s(t+\Delta_v) + n_{e,s}(t) \cdot \sigma_{e,s}) \cdot \hat{y}_{e,s}}{d(e)^2}, 0\right) \tag{7}$$

where $d(e)$ is the geographical distance between the node pair. The maximum operation ensures non-negativity of the traffic demand. This step defines a first set of traffic between all node pairs in the network. It considers the node features and the scaling based on the number of users at the nodes. However, it does not include the correct scaling between the individual service types. Therefore, in the next step, the shares between the services are adjusted. To this end, the traffic of service $s$ is normalized by the total amount of traffic generated by that service between all node pairs and at all time steps. Subsequently, the traffic is scaled with the desired share $\rho_s$ for this service type. We get

$$x''_{e,s}(t) = x'_{e,s}(t) \frac{\rho_s}{\sum_{t' \in T, e' \in E} x'_{e',s}(t')} \tag{8}$$

from which we can compute the corresponding up- and downstream demands $x''_{e,s}(t) \cdot (1 - f_s)$ and $x''_{e,s}(t) \cdot f_s$, respectively. In the next step, the upstream demand between the nodes $u$ and $v$ and the downstream demand between $v$ and $u$ of each service can be added, because they have the same direction, i. e.

$$x'''_{(u,v),s}(t) = x''_{(u,v),s}(t) \cdot (1 - f_s) + x''_{(v,u),s}(t) \cdot f_s. \tag{9}$$

In the last step the peak hour traffic, i. e. the sum of all demands maximized over all points in time, is adjusted by

$$x_{e,s}(t) = x'''_{e,s}(t) \frac{m}{\max_{t' \in T} \sum_{s' \in S, e' \in E} x'''_{e',s'}(t')} \tag{10}$$

where $m$ is the desired peak hour traffic value.

# 4 Application Example

This section provides an example of how to use the given modeling approach to create a meaningful traffic scenario and how it can be used to perform network-level studies.

## 4.1 Network and Traffic

We present an example application of the model for the US backbone network "Abilene" found in the SNDlib [3]. The network consists of 12 nodes and 15 links and is depicted in Figure 1. We describe the modeling of the network parameters and of the services and give an overview of the resulting traffic demands over one weekday.

The number of connected users at each node is derived by summing up the population of the surrounding states [14] as shown in Figure 1. As features we select *data centers* and *Internet exchange points*. Data centers are placed according to the data center locations of Google [15] and Amazon [16]. Internet exchange points are located in New York City, Chicago, Houston and Sunnyvale

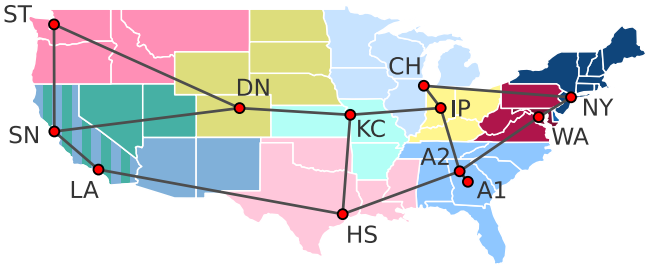


**Figure 1** US research network topology "Abilene" with user assignment.

**Table 1** Parameters for the Abilene network.

| City | Node | Users | Features | Time Offset |
|---|---|---|---|---|
| Atlanta | A1 | 1,000,000 | | 0 h |
| Atlanta | A2 | 57,302,298 | DC | 0 h |
| Chicago | CH | 37,282,134 | IX | −1 h |
| Denver | DN | 9,731,604 | | −2 h |
| Houston | HS | 39,903,893 | IX | −1 h |
| Indianapolis | IP | 22,779,616 | DC | 0 h |
| Kansas City | KC | 12,030,934 | DC | −1 h |
| Los Angeles | LA | 28,872,666 | | −3 h |
| New York City | NY | 43,665,044 | IX | 0 h |
| Sunnyvale | SN | 25,868,199 | DC, IX | −3 h |
| Seattle | ST | 14,315,955 | DC | −3 h |
| Washington, D.C. | WA | 30,799,502 | DC | 0 h |

which matches the locations of the DE-CIX [17] and AMS-IX [18] exchange points. Table 1 summarizes the network parameters. As reference timezone we choose the Eastern Time Zone. The locations of the nodes can be found in the SNDlib [3].

The modeled service types together with their overall traffic shares are based on information found in the Visual Networking Index 2016–2021 [1] and the Global Cloud Index 2016–2021 [19] issued by Cisco. The two reports contain detailed estimations of IP traffic in the United States for the years 2016 to 2021. The traffic estimated in the reports can be divided into the categories consumer, business, inter-data center and intra-data center. While the last traffic category exists only inside the data center and is not relevant for us, the first three constitute all IP traffic that also flows through wide area networks. We modeled this traffic by 10 different service types listed in Table 2.

*Internet Video* includes services like Netflix, YouTube or Amazon Video to name the three biggest providers [9]. The service *Consumer Miscellaneous* contains all consumer traffic that is not modeled by *IPTV*, *Internet Video* or *Gaming*. For business traffic we assume that all video traffic originates from *Video Conferencing* and that *Backup* and *Business Miscellaneous* have equal shares. Further we assume that 25 % of the data center traffic account for connections between data center and Internet exchange while the remaining 75 % of the traffic flow between data centers only. We further detail the inter-data center traffic by modeling two service classes, namely *DC–DC (H)* without delay requirements and *DC–DC (L)* with delay-requirements.

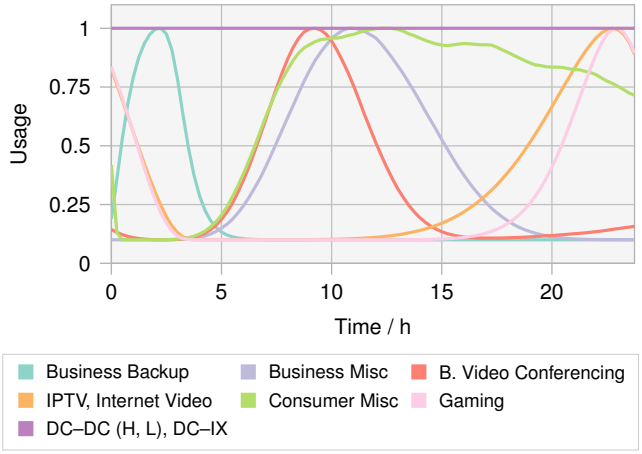


**Figure 2** Usage profiles for the modeled services.

The downstream traffic portions $f_s$ are derived from [9] in the cases of *IPTV* and *Internet Video* and from [20] for *Gaming*. For all other services we used assumptions based on own estimates. The usage profiles $p_s$ are derived from the default model of the traffic generator developed in the DISCUS project [12, 13]. In the DISCUS traffic generator the activity, w. r. t. different service types, of individual users is computed based on probability distributions. These distributions model the start time and duration of the service usage as well as the number of repetitions per day and the gap between successive service usages. For this example we use the distributions to create activity profiles for 1,000,000 users. For a service $s$, the sum of the individual activity profiles forms a usage profile $p_s$. The usage profiles of the data center services are assumed to be constant. Figure 2 shows all the usage profiles of this example. The minimum value is set to 0.1 because it is unlikely that a service is completely unused for a longer period of time.

Using the described parameters, we generated traffic demand matrices for one weekday with an interval of 15 minutes. Figure 3 shows stacked plots of the dynamic behavior of the individual services for the years 2016 and 2021. In these plots, the demands between all node pairs have been aggregated. The peak hour traffic grows from 48 Tbps in 2016 to 170 Tbps in 2021. As can be seen, the two major components are data center and video traffic. Video traffic is also the dominating part during the afternoon and night hours. Due to the four different timezones the network spans, even long after midnight the demand level is high. Figure 4 shows the fraction of delay-sensitive traffic demands for each time step. The fraction ranges from 15.0 % to 26.5 % in the year 2016 and from 16.1 % to 28.7 % in 2021. As can be seen, the share of delay-sensitive demand in 2021 will be larger than 2016 during the late morning and noon hours and also during the night hours. This increase is mainly driven by the growth of delay-sensitive video conferencing and gaming demands.

## 4.2 Delay Classes and Hardware

In order to evaluate the effects of different delay-requirements we have varied the maximum allowed delay

**Table 2** Parameters of the modeled services.

| Service | Required Features | | Downstream Portion | Share | | Delay-Sensitive |
|---|---|---|---|---|---|---|
| | Source | Target | | Year 2016 | Year 2021 | |
| **Consumer** | | | | | | |
| IPTV | | | 0.98 | 12.1 % | 8.8 % | Yes |
| Internet Video | | | 0.95 | 31.0 % | 25.1 % | No |
| Gaming | | | 0.8 | 0.7 % | 2.9 % | Yes |
| Miscellaneous | | IX | 0.75 | 6.2 % | 3.2 % | No |
| **Business** | | | | | | |
| Backup | | | 0.1 | 2.0 % | 1.2 % | No |
| Video Conferencing | | | 0.5 | 4.1 % | 5.1 % | Yes |
| Miscellaneous | | IX | 0.75 | 2.0 % | 1.2 % | No |
| **Data Center** | | | | | | |
| DC–DC (H) | DC | DC | 0.5 | 27.2 % | 34.1 % | No |
| DC–DC (L) | DC | DC | 0.5 | 4.2 % | 5.3 % | Yes |
| DC–IX | DC | IX | 0.5 | 10.5 % | 13.1 % | No |

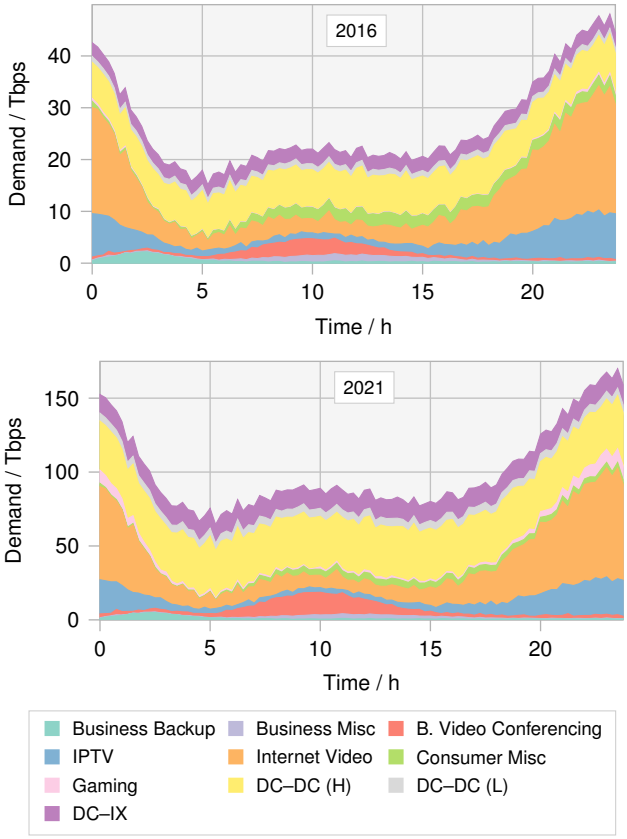


**Figure 3** Stacked plots of the overall traffic demand sums in the years 2016 and 2021.

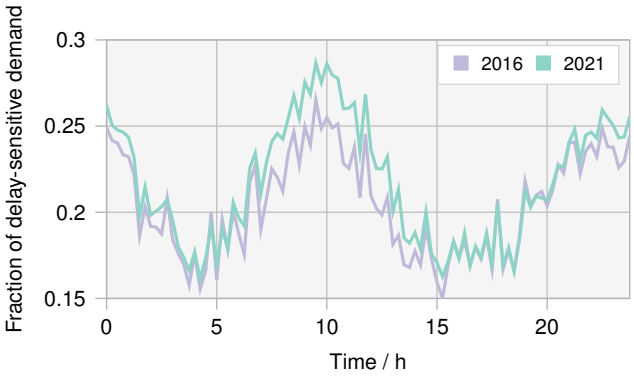


**Figure 4** Fraction of delay-sensitive demand.

**Table 3** Lowest delay class possible between node pairs.

| | A1 | A2 | CH | DN | HS | IP | KC | LA | NY | SN | ST | WA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | – | 5 | 5 | 15 | 10 | 5 | 10 | 25 | 10 | 25 | 25 | 10 |
| A2 | 5 | – | 5 | 15 | 10 | 5 | 10 | 25 | 10 | 25 | 25 | 5 |
| CH | 5 | 5 | – | 10 | 10 | 5 | 10 | 25 | 10 | 25 | 25 | 10 |
| DN | 15 | 15 | 10 | – | 10 | 10 | 5 | 15 | 25 | 10 | 10 | 25 |
| HS | 10 | 10 | 10 | 10 | – | 10 | 10 | 15 | 15 | 15 | 25 | 10 |
| IP | 5 | 5 | 5 | 10 | 10 | – | 5 | 25 | 10 | 25 | 25 | 10 |
| KC | 10 | 10 | 10 | 5 | 10 | 5 | – | 15 | 15 | 15 | 15 | 15 |
| LA | 25 | 25 | 25 | 15 | 15 | 25 | 15 | – | 25 | 5 | 10 | 25 |
| NY | 10 | 10 | 10 | 25 | 15 | 10 | 15 | 25 | – | 25 | 25 | 5 |
| SN | 25 | 25 | 25 | 10 | 15 | 25 | 15 | 5 | 25 | – | 10 | 25 |
| ST | 25 | 25 | 25 | 10 | 25 | 25 | 15 | 10 | 25 | 10 | – | 25 |
| WA | 10 | 5 | 10 | 25 | 10 | 10 | 15 | 25 | 5 | 25 | 25 | – |

of delay-sensitive traffic classes. Based purely on propagation delay, which is the most prominent contributor in wide area networks, we first determined which demands could be served at which delay due to geographical constraints. Table 3 shows which node pairs can be served with delay classes guaranteeing 5, 10, 15 and 25 ms.

Using these values we created four different sets of dynamic traffic matrices. Our demand set for 25 ms requires all delay-sensitive demands to be in the 25 ms-class. Subsequent sets replace this constraint with the next lower class where possible meaning that e. g. the 15 ms-set requires all demands to have 15 ms if possible and 25 ms otherwise. The 5 ms-set requires all demands to be in the lowest-possible of these four classes, corresponding di-

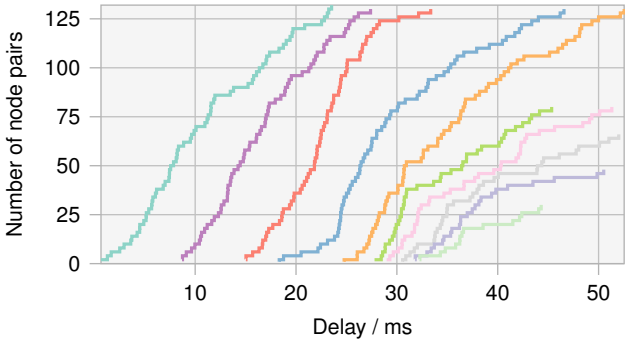


**Figure 5** Distribution of the propagation delay for the 10 shortest paths between each node pair. The path index increases from left to right, i. e. the green curve on the left corresponds to the shortest path.

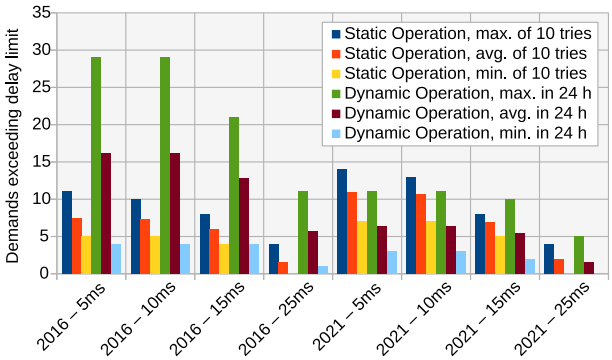


**Figure 6** Number of demands exceeding delay limit if available resources are minimal according to the delay-unaware optimization heuristic.

rectly to the entries in Table 3.

Figure 5 shows a theoretical analysis of the degrees of freedom in routing permissible by different delay restrictions. For a 5 ms delay limit there are only 22 node pairs available and only the shortest path (green on the left) is able to keep this limit. For the most lenient limit of 25 ms it can be observed that almost all node pairs (98.5 %) have at least one alternate path (purple) which can keep the delay limit, while some few paths have up to four alternate paths.

To show the difference between static and dynamic traffic we have derived static traffic matrices for each scenario as well. These were created by determining the maximum traffic value for each node pair throughout the sequence of dynamic matrices for the respective scenario. Since our example time frame stretches from the year 2016 to 2021 we chose to use hardware parameters which have just become commercially viable in 2016 and will still be relevant in 2021. Therefore, we assume our transponders to be operating at a data rate of 200 Gbps utilizing PDM-8QAM and router line cards supporting 4 ports of this capacity. At a transparent reach of about 1800 km [21] these devices can cover all the single-hop distances in the network except for the link between Houston and Los Angeles which will require regeneration.

### 4.3 Dimensioning and Reconfiguration

We have developed several optimization heuristic approaches for the dynamic reconfiguration of multi-layer networks in previous works [22, 23] which can also be applied in network dimensioning. The objective of these approaches is to determine a configuration of network hardware resources and corresponding routing in the face of dynamically changing traffic demands considering given optimization criteria. In the past we had considered energy-efficiency through deactivation of unneeded hardware and resource efficiency in terms of installed hardware required for the given traffic. To demonstrate the effects of heterogeneous and dynamic traffic behavior we have conducted a number of simulation studies based on assumptions in the previous sections. For this study we have used our virtual topology-centric optimization heuristic [23] in two versions. One is focused solely on resource optimization and does not consider any delay requirements and serves as a baseline for the second version which additionally takes delay constraints into account. We will evaluate their effectiveness in dimensioning and reconfiguration by the number of line cards required for the entire network.

## 5 Simulation Results

Using the delay-unaware heuristic, that optimizes resource usage only, we determined the hardware requirements for all 8 scenarios for two modes of network operation. For the regular static operation we determined resource requirements for the static matrices. In the dynamic mode, the network is reconfigured every hour to use the least amount of resources for the present matrix in the dynamic traffic set. In theory, a resource optimization will route traffic on shortest paths and only divert to longer ones if grooming of traffic allows to save otherwise underutilized interfaces on more direct routes. Since routing all traffic on shortest paths will automatically achieve the least delay possible, we first ran a resource optimization that is purposefully unaware of the delay constraints to estimate the effects of grooming. In order to control for statistical effects we solved each static matrix 10 times. The dynamic matrices were only done once and therefore the corresponding results should only be considered as an estimate.

Figure 6 shows that for the static case only few of the 132 delay-sensitive traffic demands have not met their delay requirements. The dynamic case using the periodic network reconfiguration shows a much larger number of excessive delays for the 2016 scenarios due to the fact that it vastly reduces the number of active line cards during the hours of lower traffic load thereby reducing the number of direct routes and thus increasing the average path length by 6 % and the maximum path length by 21.3 %. For the 2021 scenarios the average traffic load is much higher which necessitates using about three times as many line cards. Since for this case there is a significant load on almost all possible links, the optimization algorithm finds only few meaningful opportunities to provide grooming of traffic on longer alternate routes and the results therefore almost match the

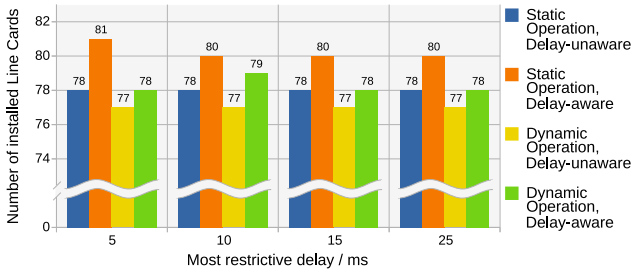


**Figure 7** Minimal number of installed line cards required for the 2016 scenarios for different operation modes.

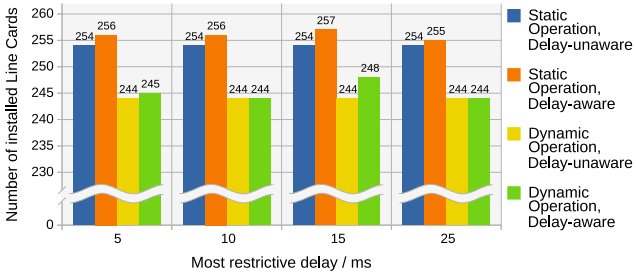


**Figure 8** Minimal number of installed line cards required for the 2021 scenarios for different operation modes.

static case.

We repeated all of the experiments above using the second, delay-aware version of our optimization heuristic. Since avoiding delay excursions reduces the potential for grooming traffic onto longer paths, the resource requirements could be expected to be significantly higher than in the previous experiments. Figure 7 and Figure 8 show a comparison of the lowest number of line cards the algorithms have found to be required for each scenario. We found that even the most stringent delay requirements can be fulfilled by adding at most three additional line cards to the absolute resource minimum corresponding to a relative increase of at most 3.85 % for 2016 scenarios and 1.18 % for the 2021 scenarios. Surprisingly, the intuitive correlation between lowered delay constraints and increased number of required line cards is not visible in the data shown, although it has been observed in other studies with higher peak traffic. However, the sample size and numerical differences in the present data are too small to draw conclusions on the cause of this effect. On the other hand, the comparison between static and dynamic cases shows that the dynamic operation can save at least between 1.3 % and 4.69 % of line cards in the given scenarios. This is due to the fact that traffic peaks between node pairs do not always occur simultaneously allowing some line cards to be repurposed during network operation where for the static case two line cards are needed.

Figure 9 and Figure 10 show the results of the dynamic operation experiments in more detail. For the course of a simulated time of 24 hours the network configuration is adjusted every hour to the present traffic situation in order to deactivate the most number of line cards possible making these graphs follow the shape of the traffic graphs in

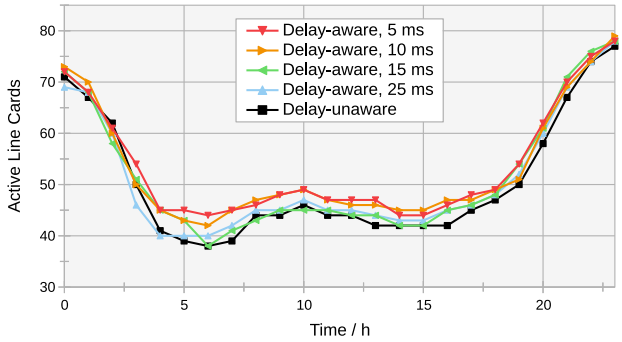


**Figure 9** Number of active line cards required in the course of the day for the 2016 scenario.

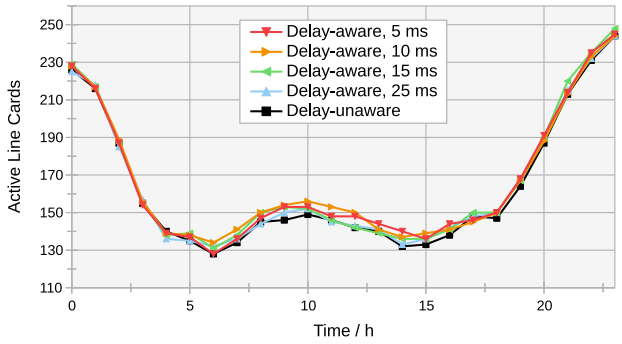


**Figure 10** Number of active line cards required in the course of the day for the 2021 scenario.

Figure 3. This approach can be used in combination with hardware sleep modes to reduce operational expenditure in terms of energy cost. For all scenarios the behavior is similar and the average number of line cards that have to remain active ranges between 65 % and 68 % of the installed line cards. During the times of low network load honoring more demanding delay constraints consistently requires a larger number of active line cards. For the 2016 case each next lower delay class requires about 1 % additional resources compared to the resource-unaware case while the relative difference in the 2021 case always stays below 1 %.

In summary of these results it can be stated for the given scenario that while offering increasingly low delay classes will require additional resources, the amount is relatively low. Even in the dynamic case the increase in line cards required to be active in order to avoid missing the delay limits, especially at times of low utilization, has only a small negative impact considering the observed maximum number of delay excursions for the baseline case.

# 6 Conclusion

In this work we have presented requirements for a traffic generator supporting dynamic demands. We have found that to our knowledge there is no modeling approach in literature which provides the necessary layer of abstraction with the required level of precision. We have suggested a modeling approach which allows to create traffic scenar-

ios based on population figures, usage profiles and traffic volume extrapolations readily found in literature. We presented an application example validating the capabilities of this approach by creating traffic scenarios matching current traffic predictions and demonstrating how they can be applied to network-level studies on resource dimensioning and dynamic operation.

# 7 Acknowledgments

This work has been performed in the framework of the CELTIC EUREKA project SENDATE-TANDEM (Project ID C2015/3-2), and it is partly funded by the German BMBF (Project ID 16KIS0458). The authors alone are responsible for the content of the paper.

# 8 References

[1] Cisco. Visual Networking Index: Forecast and Methodology, 2016–2021. Online, June 2017.

[2] S. Talwar, D. Choudhury, K. Dimou, E. Aryafar, B. Bangerter, and K. Stewart. Enabling Technologies and Architectures for 5G Wireless. In *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, pages 1–4, June 2014.

[3] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly. SNDlib 1.0–Survivable Network Design Library. In *Proceedings of the 3rd International Network Optimization Conference (INOC 2007), Spa, Belgium*, April 2007.

[4] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon. Providing Public Intradomain Traffic Matrices to the Research Community. *SIGCOMM Comput. Commun. Rev.*, 36(1):83–86, January 2006.

[5] P. Tune and M. Roughan. Internet Traffic Matrices: A Primer. *Recent Advances in Networking*, 1:1–56, 2013.

[6] J.P. Kowalski and B. Warfield. Modelling Traffic Demand between Nodes in a Telecommunications Network. In *ATNAC'95*, 1995.

[7] A. Dwivedi and R.E. Wagner. Traffic Model for USA long-distance Optical Network. In *Optical Fiber Communication Conference, 2000*, volume 1, pages 156–158. IEEE, 2000.

[8] V. Erramill, M. Crovella, and N. Taft. An Independent-Connection Model for Traffic Matrices. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 251–256. ACM, 2006.

[9] Sandvine Incorporated ULC. Global Internet Phenomena Report 2016. Online, June 2016.

[10] A. Nucci, A. Sridharan, and N. Taft. The Problem of synthetically Generating IP Traffic Matrices: Initial Recommendations. *ACM SIGCOMM Computer Communication Review*, 35(3):19–32, 2005.

[11] DE-CIX Management GmbH. DE-CIX Frankfurt Statistics. Online. URL `https://de-cix.net/en/locations/germany/frankfurt/statistics`. Accessed: 2017-05-19.

[12] M. Ruffini. trafficGen – GitHub. URL `https://github.com/ruffinim/trafficGen`. Accessed: 2018-04-06.

[13] M. Ruffini, L. Wosinska, M. Achouche, J. Chen, N. Doran, F. Farjady, J. Montalvo, P. Ossieur, B. O'Sullivan, N. Parsons, et al. DISCUS: An end-to-end Solution for ubiquitous Broadband Optical Access. *IEEE Communications Magazine*, 52(2):S24–S32, 2014.

[14] US Census Bureau. National, State, and Puerto Rico Commonwealth Totals Datasets: Population, population change, and estimated components of population change: April 1, 2010 to July 1, 2017. Online. URL `https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/national/totals/nst-est2017-alldata.csv`. Accessed: 2018-03-14.

[15] Google LLC. Data center locations – Data Centers – Google. URL `https://www.google.com/about/datacenters/inside/locations/`. Accessed: 2018-03-15.

[16] Amazon Web Services, Inc. Global Cloud Infrastructure | Regions and Availability Zones | AWS. URL `https://aws.amazon.com/about-aws/global-infrastructure/`. Accessed: 2018-03-15.

[17] DE-CIX Management GmbH. Locations – DE-CIX. URL `https://www.de-cix.net/en/locations/`. Accessed: 2018-03-15.

[18] AMS-IX USA Inc. AMS-IX US Internet Exchange Home. URL `https://us.ams-ix.net/`. Accessed: 2018-03-15.

[19] Cisco. Global Cloud Index: Forecast and Methodology, 2016–2021. Online, June 2017.

[20] X. Wang, T. Kwon, Y. Choi, M. Chen, and Y. Zhang. Characterizing the Gaming Traffic of World of Warcraft: From Game Scenarios to Network Access Technologies. *IEEE Network*, 26(1), 2012.

[21] B. Lavigne, M. Lefrançois S. Weisser, R. Peruta, G. A. Azzini, and L. Suberini. Real-time 200 Gb/s 8-QAM Transmission over a 1800-km long SSMF-based System using add/drop 50 GHz-wide Filters. In *2016 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, March 2016.

[22] U. Bauknecht and F. Feller. Dynamic Resource Operation and Power Model for IP-over-WSON Networks. In *Proceedings of the 19th Open European Summer School and IFIP TC6.6 Workshop (EUNICE 2013)*, 2013.

[23] U. Bauknecht. Resource Efficiency and Latency in Dynamic IP-over-WSON Networks utilizing Flexrate Transponders. In *Photonic Networks; 18. ITG-Symposium; Proceedings of*, pages 1–6. VDE, 2017.