

Guidelines for Connection-Level Performance Simulation of Optical Networks

Dominic A. Schupke, Svetoslav Duhovnikov and Clara Meusburger

Abstract— Simulation of optical networks that set up and tear down connections dynamically involves defining the network and traffic model as well as configuring the simulator itself, to investigate a researched method, for example, a novel routing scheme. Because of interdependences in the overall constellation, a coordinated setting is necessary to obtain meaningful performance results such as estimates on the blocking probability. This article develops a framework for connection-level simulation along with several suggestions on further simulation enhancements.

Index Terms—Simulation, optical networks, performance evaluation, confidence of results

I. INTRODUCTION

ONE of the appealing options of future optical networks is the operation with dynamic connections. The connections could be opaque or transparent lightpaths. We think of “dynamic” as changing over time and these time changes occur within the order of weeks or shorter. Since we are dealing with future, not yet existing systems, simulation of such networks is a powerful method to evaluate their performance. While early research has often only been able to assume best-guess assumptions, progressed research can now rely on much more detail and flexibility for simulation.

How to obtain meaningful results in the area of simulation is an often discussed point [1], leading to some criticism. Reactions to published work come even down to statements like “Well I think the math should be all right, but what about the assumptions?” This article aims to give recommendations to conduct connection-level performance simulation of optical networks. Following the recommendations, especially adopting well chosen and plausible assumptions, increases the value of the work. As a result, it is more likely, firstly, that corresponding publications are rated high by reviewers and are frequently cited, and, secondly, that results are accepted as useful knowledge, especially by industry experts. Clearly, we cannot model and simulate any minute detail, however, it is

important that we mind principal points.

Simulation can assess novel methods, such as the performance that a planning method yields for operation, the comparative performance of different Routing and Wavelength Assignment (RWA) algorithms, the performance of algorithms that place network elements (such as waveband switches), or the performance of recovery mechanisms. Many of the raised points are also relevant for other network types, such as Optical Burst Switching (OBS) networks as investigated in [2] for the dependence of network performance on network dimensioning.

An underpinning paper on network simulation is [1]. The book [3] presents several recent study results on modeling and simulation of communication networks, such as multilayer modeling. Network simulation validation is discussed in [4], illustrating the related issues around TCP simulation.

As a leitmotiv for selecting a scenario, one should give enough detail such that the simulation can be reproduced by others. This is encouraged since many research institutions have a simulation environment in place and thus can build upon these details for further studies. A scenario consists of the network model, the traffic model, the simulator, and the investigated method. The next two sections deal with the network and the traffic models, the following section goes into more detail of simulators, allowing for generic recommendations on investigations of different methods. The last two sections give an outlook and conclusion of the paper.

II. NETWORK MODELS

In the real world, demand is carried over a planned network, i.e., operators dimension the capacities and select equipment to carry a given or predicted demand. This planning is necessary because, firstly, optical networks involve large investments of the operator and, secondly, underprovisioning (even at few points) can cause service rejection, and, thus, revenue losses.

Due to this fact alone, simulation should use network models that are designed for their subsequent operation. To illustrate the effects that can occur, consider blocking probability simulation of the well-known NSFNET in Figure 1, which has 14 nodes and 21 topology links. In this article, we assume transparent WDM networks (i.e., no wavelength conversion) and, unless mentioned otherwise, physical length serves as routing metric.

D. A. Schupke is with Nokia Siemens Networks, St.-Martin-Str. 53, 81669 Munich, Germany, Email: schupke@ieee.org

S. Duhovnikov is with Nokia Siemens Networks, St.-Martin-Str. 53, 81669 Munich, Germany, Email svetoslav.duhovnikov@nsn.com

C. Meusburger is with Nokia Siemens Networks, St.-Martin-Str. 53, 81669 Munich, Germany, and also with the Technische Universität München, 80333 Munich, Germany, Email: clara.meusburger.ext@nsn.com.

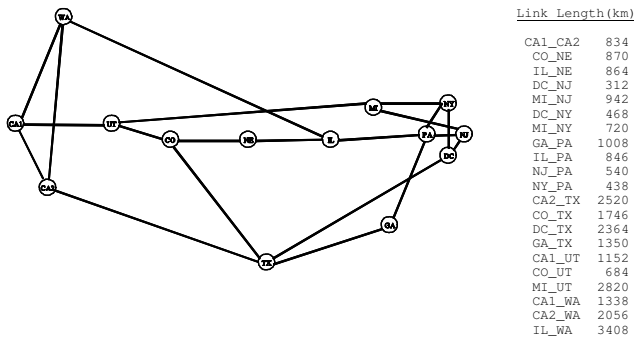


Fig. 1: This US topology is based on the National Science Foundation Network (NSFNET) which can be found, e.g., in [5].

In a simple approach, we could firstly assign the same capacity to every topology link, i.e., a “flat network” where every topology link has the same number fiber links, and secondly assume every nodepair to have the same demand, i.e., a homogeneous “any-to-any” relationship. Shortest path routing yields the link-load distribution depicted in Figure 2. The number of shortest paths on topology link IL_PA is at least one third and up to twelve times higher than on other links. Consequently, link IL_PA likely dominates the overall blocking probability.

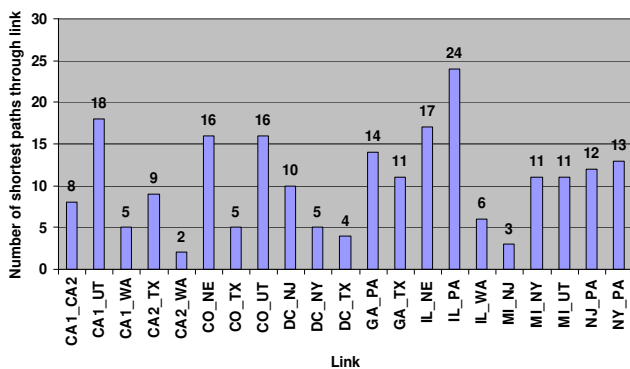


Fig. 2: The number of length-based shortest paths traversing a topology link in the US topology of Figure 1 for an any-to-any demand.

To obtain further notion, we turn aside to a static planning problem that is equivalent to the ‘linear dimensioning’ [2] for dynamic networks. For the sake of illustration, assume that fiber links can be set up with granularity of eight wavelengths and the network is offered an any-to-any demand of one unit. In reality we would expect higher values, but if we scale the numbers by, say, five we attain a realistic region and can draw analogous conclusions. In a flat network, we need three fiber links (summing to 24 wavelengths) for each topology link and thus 504 wavelengths overall in the network, to achieve unblocked routing. A linear dimensioning, which rounds up to the next multiple of eight for each topology link, needs only 296 wavelengths overall in the network, using shortest path routing. Although traffic generation itself also impacts the results (see Section III), firstly, this comparison underlines the sake of conscious network planning. Secondly, it also shows that a flat network with one fiber link or two fiber links per

topology link effects blocking and inhomogeneous dimensioning. Out of the 21 topology links, 13 in the first case (one fiber link) and 3 in the second case (two fiber links) links account for blocking, see Figure 2. With three and more fiber links the network becomes overdimensioned.

Note that even though relative performance of different routing schemes is of interest, generally the performance of a specific routing scheme depends on the working area.

This behavior is also reflected in the dynamic operation scenario. Figure 3 depicts simulation results of the estimated blocking probability for Shortest Transparent Path (STP) and Shortest Available Transparent Path (SATP) routing. STP routing makes only use of the shortest path and assigns the first-fitting available wavelength for a connection request between a nodepair. SATP routing searches the shortest possible path for which a continuous wavelength is available (the first-fitting available wavelength is then used). If a continuous wavelength is not available, for STP on the shortest path and for SATP on any path, we encounter a blocking-event.

We assume an any-to-any demand that is scaled by the value on the x-axis (called demand scaling factor). Both interarrival times and holding times per demand-unit are negative exponentially distributed and have means of 10 time units and 1 time unit, respectively. For STP, the results show that linear dimensioning is superior to the flat network realizations with one or two fiber links per topology link. It is only inferior for the overdimensioned case of three fiber links per topology link. For linear dimensioning, SATP outperforms STP.

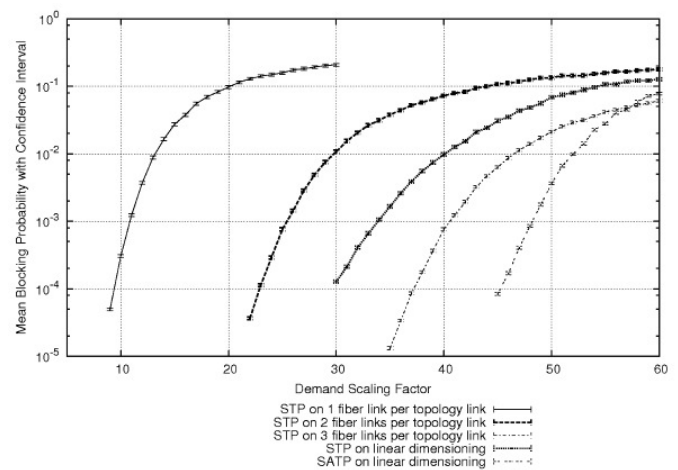


Fig. 3: The simulated mean blocking probability with relative confidence interval (CI) of 5% at 95% confidence level in the NSFNET for Shortest Transparent Path (STP) and Shortest Available Transparent Path (SATP) routing, on different dimensionings. We assume a transparent network with 8 wavelengths per fiber link.

As another aspect, we should avoid large deviations between the routing used in network planning and the routing used in operations. The complexity of planning problems often mandates that routing methods are simpler than those ones used in operations (e.g., using fewer paths), but we should aim at minimizing obvious incompatibilities. One pitfall example is

that both network planning and operations use shortest path routing, however, multiple shortest paths exist between nodepairs. This case often occurs for hop-based shortest path routing, for example, in the NSFNET 21 out of its 91 nodepairs have two or more hop-minimal paths. If the network planning procedure dimensions the links assuming shortest paths that are different from those shortest paths used in operations, we can again run into high blocking values. This happens on links where the dimensioned capacity is insufficient, while on other links excessive capacity is available.

For meaningful comparative results, e.g., comparing RWA algorithms, the dependence of the operation on the given network can also come to the fore. Assume a simulative comparison shows that a novel algorithm B performs better than a known algorithm A for an unplanned (e.g., flat) network. It seems unfair to generally conclude that B is better than A, if the algorithms are operated on a dimensioned network, changing the mutual performance relation. Therefore, we recommend to determine first the traffic model and perform network planning, and then to simulate the network operation.

III. TRAFFIC MODELS

Once the network is modeled, the traffic offered to the network can be defined. Finding a representative traffic model seems to be the hardest part in the simulation of dynamic optical networks. While many realistic traffic models for services on higher layers are at hand, it remains currently only a guess when and where connection demand in future optical networks occurs. Again, the choice of a specific traffic type should be made plausible.

As for the connection arrival process, the bulk of literature resorts to a Poisson process that is directly coupled to a negative exponentially distributed interarrival time, which matches telephone traffic well. A direct transformation to traffic in optical networks remains apparently unreasoned, since the future user behavior is yet unknown for optical networks. It turns out, however, that a Poisson arrival process is best practice at the moment, because of its memory-less property. In effect, we can model that an operator does not know when a connection will be set up. Another reason for researchers favoring the Poisson arrival process lies in the long experience in its characteristics and in the possibility to achieve results comparable with studies from various authors.

As for the connection holding time, literature mainly adopts a negative exponential distribution. Similar to the interarrival time, this assumption is often a best guess for dynamic traffic, e.g., where an operator does not know when a connection will be torn down.

A dynamic traffic matrix is a common way to express the offered traffic between nodepairs. For example, a triangle matrix with equal entries can render a bidirectional any-to-any demand. If both interarrival time and holding time are negative exponentially distributed with means $1/\lambda_{i,j}$ and $1/\mu_{i,j}$,

respectively, the entries between nodes i and j of the dynamic traffic matrix are given as traffic volume values $\lambda_{i,j}/\mu_{i,j}$ measured in Erlang. In other cases, we may need more complex, less intuitive characterizations, thus, as for the planning of the network, it is important to accurately describe the model of offered traffic.

Another holding-time operation that merely matches today's situation, however, assumes that connections are never torn down. Together with over-time arrivals we obtain traffic of type 'incremental' that offers new challenges in the simulation modeling (see Section Outlook) and performance assessment. In networks with a given capacity, simulating a steady-state blocking probability of incremental traffic becomes meaningless, since this probability tends to one. A doable performance assessment option is to measure after how many connection set-ups a first blocking occurs for random arrival sequences starting on empty networks. This can be depicted in a probability density function (pdf) as in Figure 4, or otherwise using mean values plus confidence intervals (defined in the Presentation of Results Section below). For the 10,000 simulated sequences and for a 95% confidence level, the mean plus confidence interval is 225 ± 0.5 and 522 ± 0.5 for STP and SATP, respectively.

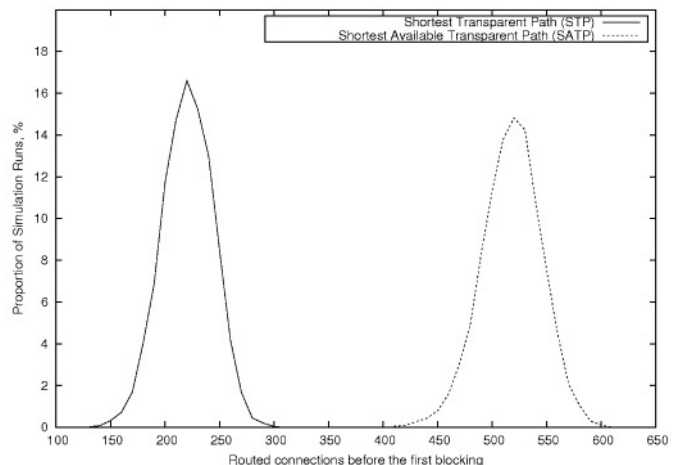


Fig. 4: Simulated probability density function (pdf) for the number of routed connections before the first blocked request occurs, comparing the two algorithms STP and SATP.

In the recent years, literature puts forward the view of “What should the user traffic look like, such that we can exploit the network resources best?” aiming at the definition of novel service types, see e.g. [6,7]. One of the possibilities is scheduled traffic that can be accurately characterized by the definition of the service. An efficient utilization of the service can be effected by a suitable tariff.

Finally, the traffic can differ according to service types. While our presentation assumes a single unprotected connection demand, it is often interesting to generalize to other ones, even co-existing types. One alternative is that multiple connections arrive at a time, to model bundled requests or Virtual Private Networks (VPNs). Unprotected connections with diversity constraints (to allow for protection on higher layers) and protected connections are other service classifiers.

For instance, higher-layer protected VPNs may be inadequately modeled by single unprotected connections without diversity constraints. In the following, we sketch how realistic modeling of such a VPN service can augment the complexity of the simulation. Firstly, especially as the VPN service requires diversity among the requested connections, a dedicated VPN routing process with blocking rules is needed. Secondly, since VPNs arrive and depart at a time, an own arrival/departure process for a set of connections is needed. Thirdly, the simulator needs to generate realistic VPN structures (i.e., selections of nodes and virtual links) and model the modification process of existing VPNs (i.e., enlarge or shrink VPNs).

IV. SIMULATOR

This section deals with the characteristics of the widely-adopted steady-state simulation, i.e., a simulation of the state when time tends to infinity. The Outlook Section pursues alternatives to steady-state models.

To model randomly occurring events in the network, simulators rely on a Pseudo-Random Number Generator (PRNG). As pointed out in [1], PRNGs should have cycles that are long enough to avoid PRNG-induced correlations during the simulation. These correlations can produce misleading results, because the randomness of the modeled stochastic process is insufficient. While PRNGs with cycles in the order of 232 may be sufficient in several network-level simulations, using higher orders puts us on the safe side. In any case, the PRNG used during the simulation and its configuration (e.g., seeds) should be mentioned. For the simulations in this article, we employ the PRNG “Mersenne Twister” (cited in [1]) with the cycle of $2^{19937}-1$ and seed set to 2326.

At the beginning of a simulation, networks carrying dynamic traffic are typically in a non-representative state, e.g., networks are initially empty and then filled up with traffic until a stable operation point is reached. Additionally, an insufficient number of events has yet been simulated, biasing the simulated outcome. Figure 5 exemplifies this where blocking events are at the beginning rare because of a non-filled network. Including this “warm-up period” to estimate steady state can cause a significant bias in the final results [1]. Reports on simulation should therefore state how the warm-up period is handled, e.g., when the end of the warm-up period is determined and that this warm-up period is excluded from the calculation of the results.

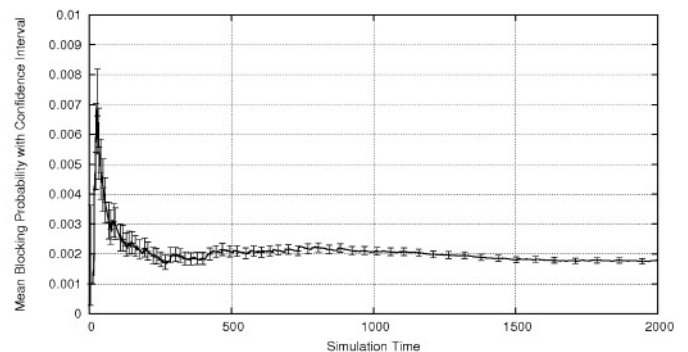


Fig. 5: Simulated mean blocking probability plus confidence interval (CI) over simulation time for Shortest Available Transparent Path in the linear-dimensional NSFNET at demand scaling of 49.

A simple, yet cumbersome way to determine the warm-up period is to inspect it visually in the data, e.g., in Figure 5 we could say that after 500 simulation time units the warm-up period expired. A common test for automated simulation is to determine the end of warm-up after the interpolation of result values crossed the mean of these values k times. On the one hand, to keep the overall simulation time short, we aim to minimize the warm-up period, hence we aim to set the lowest k . On the other hand, k has to be sufficiently large to cancel out biases in the results. In our investigations this test is less generic for network simulation, since k highly depends on the actual network instance (we saw practical values of $k=400$ to 2500). Therefore, we regard automating tests for determining the warm-up period in network simulations as an area of future research.

Simulations should run until statistically significant results are computed. Advances in computer technology have made running time less an issue, but still, complex network simulations are exposed to be stopped too early. If users cannot wait until the simulator has proceeded enough and thus it cannot guarantee credible results, why should we show and discuss simulation results at all? A widely adopted method is to check the confidence interval (CI) of each measured parameter [1]. The CI at a given confidence level (CL) states that the measured parameter, e.g., a mean blocking probability, lies with probability of CL within this CI. Typical CL values are 95% and 97.5%. During the simulation, as more statistical data is gathered, the CI reduces. The CI can be checked after a fixed-length simulation, to see if the interval is small enough, or it can be conveniently used as stopping criterion in simulation runs for a given target CI. The error bars in Figure 5 depict how the CI narrows over simulation time.

In general, simulators should be validated by comparing results at selected operation points with results from independently implemented simulators or with analytical results (such as M/M/1 and M/G/1 formulas). This helps find errors in models and simulators, and convince that the simulation produces credible results. In a further step, models, simulators, and results can even be made freely available to others, easing this kind of comparisons.

V. PRESENTATION OF RESULTS

A convincing presentation of results follows modeling and simulation work. Simulation results should discuss the confidence of the results [1]. A common way is to include confidence intervals in simulation diagrams (see Figures 3 and 6) or some statement in the text about it.

Simulation of dynamic networks typically estimates the overall network blocking probability for a given offered traffic. Because of higher event rates, operation at high blocking values (e.g., 5% and higher) are computed faster until confidence requirements are reached, and it appears that often these high-blocking regions are shown in simulation results. For regular network operation, we may question why high blocking values are of interest. For instance, 10% blocking probability means that every tenth connection request (of say 10 Gbit/s) will be rejected leading into large loss of revenue for operators of corresponding networks. Thus, it is best to simulate and discuss only those values that are of interest. A viable range is to simulate blocking values from 10^{-5} to 10^{-2} for regular operation. In this range the aspect of blocking fairness among nodepairs is also a less important issue, since blocking events are rare in total. Computing blocking probabilities below 10^{-5} involve excessive simulation times, but this extremely low blocking region can often be ignored (if we expect on average 10 connections a day, we encounter blocking on average every thousand days or more). In non-regular situations, such as high networks loads after major failure events, high blocking levels appear and if these situations need inspection it is of course feasible to simulate them.

Another effect that can cause misleading interpretations is shown in Figure 6. Assume we had only simulated the region of a demand scaling exceeding 70. Then we can conclude that STP is comparable with SATP, however, we can only say this for the high-blocking region. In the regular operation region of 10^{-2} and below, see Figure 3, the comparative performance differs significantly. A similar effect holds for comparing STP on 3 fiber links per topology link and SATP on linear dimensioning, where the relative performance swaps at a blocking level of 5% (demand scaling around 57), see Figure 3. Hence, conclusions from comparing algorithms are dependent on operation points and these operation points have to be well selected to match the intended application.

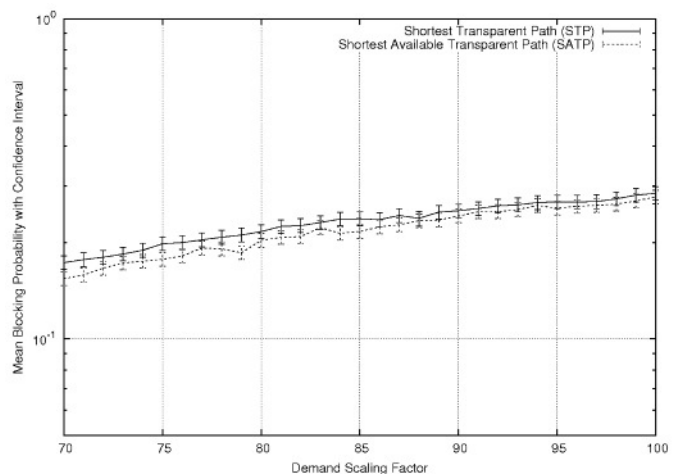


Fig. 6: Continuation of the STP and SATP curves in Figure 3, for linear dimensioning and higher demand scaling factors.

The presentation of the results should also show the per-link average load and the per-link capacity, to see whether network operation matches the dimensioning.

Simulation helps to find the behavior of the investigated method causing the computed results. Concluding beyond “algorithm B has lower blocking values than algorithm A” is of virtue, for example, if further detail (illustrated by the link loads or path lengths) supports why algorithm B performs better than A.

Ultimately, simulation results can be a fruitful basis for finding analytic expressions (bounds and approximations) on the investigated parameters. If analytic expressions are present, they can be added to diagrams and tables showing the simulation results.

Besides blocking probability, other evaluation metrics such as latency are of interest as well. Latency in packet or burst switching networks involves simulation of queues, models, processing, and propagation delays. In connection switching networks, latencies become an issue on the control plane level, involving simulation of control messages. For example, simulating the recovery time of restorable connections requires inclusion of the signaling latency after failure events.

VI. OUTLOOK

This section extends the previous discussion by providing ideas for enhancing current simulation modes and for future research topics.

As more computer processing capacity is becoming available, we enter a new age of simulation, coming even closer to real network operations. In effect, we are able to simulate multiple interrelating processes simultaneously, beyond connection set-up and tear-down.

One obvious addition to simulating dynamic connections is the concurrent failure process. Failures, such as fiber cuts, cause decrease of available capacity that can result in irregular connections or delayed set-up for new connections arriving during the repair of the failure. Adding the failure process also

allows for analysis of recovery mechanisms, e.g., their post-failure restorability (especially in presence of multiple failures).

Another hot topic is the re-optimization of a network (e.g., in maintenance windows), often referred to as part of traffic engineering. The re-optimization process aims to free capacity, e.g., we can reroute connections from long paths to shorter ones that are available by meanwhile torn-down connections. The freed capacity can affect the connection routing of future demands.

Another enhancement is the capacity upgrade of the network, especially for incremental-type traffic. Extending the network by fiber links is an example of capacity upgrade. Figure 7 sketches the network load over time for a network with constant network capacity in a typical steady-state simulation of arriving and departing connections.

Steady-State Model

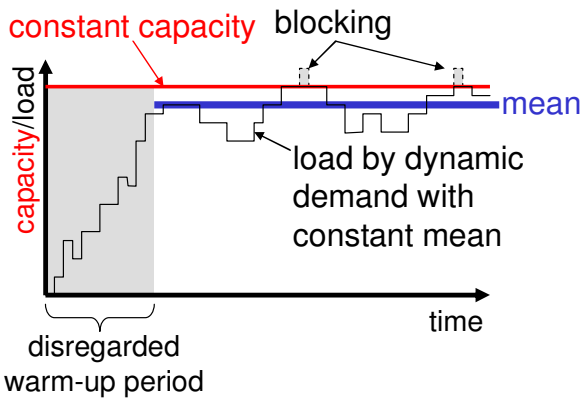


Fig. 7: In the steady-state model, the mean demand and the capacity is constant in the network.

Particularly because of the traffic demand in Internet backbones, however, today we rather see a phenomenon as outlined in Figure 8. The network load grows on average over time and the network adapts its capacity in response to this growth. In the special, yet realistic case, the network load is purely incremental (no tear-downs) and capacity adapts to completely avoid blocking. Modeling the capacity upgrade process hence becomes an appealing option to model operation of current and nearer-future networks.

Upgrading Model

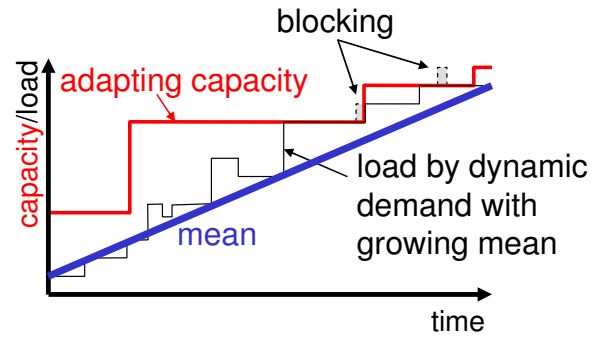


Fig. 8: In the upgrading model, demand and capacity is on average growing in the network.

Up to now we deal with simulating one network instance. While it has become common practice to investigate several representative network instances, ideally a simulator could even vary internally the dimensioned networks and the corresponding traffic demands. This way, conclusions on simulations become independent on the particularities of one network instance or few ones. The variations involve generating a large number of network topologies and demands, to obtain confident results. Along with other simulated parameters, the corresponding long simulation runs may thus still be too challenging yet. However, generating random optical topologies (such as in [8]) can soon be conceivable for simulation.

Finally, we regard sensitivity analysis as an interesting future topic. Results become more conclusive if variations to demand values, traffic characteristics, network scales, routing strategies, and technology limitations are considered in the parameterization of the simulation studies.

VII. CONCLUSION

Simulating the realistic operation of connection-level optical networks means (i) that the network is well-planned and the operation simulation relies on the necessary details from network planning, (ii) that the traffic model is completely described (volumes and time characteristics), and (iii) that the simulator ensures randomness, excludes warm-up biases, and computes until confident results are obtainable. The presentation of the results should state confidence measures and show only relevant operation ranges for any data that is necessary to understand the behavior of the investigated method. To provide even more realistic modeling, we expect that connection-level simulators will commonly be enriched by further simulatable processes, such as failures, re-optimizations, and capacity upgrading.

ACKNOWLEDGMENT

The authors thank the participants of Work Package 5 of the EU-project NOBEL for the invaluable discussions on optical

network modeling and simulation. They also thank Marco Hoffmann his comments on the manuscript.

REFERENCES

- [1] K. Pawlikowski, H.-D.J. Jeong and J.-S. Ruth Lee. On Credibility of Simulation Studies Of Telecommunication Networks. IEEE Communications Magazine, January 2002, 132-139.
- [2] Gauger, C.M.; Köhn, M.; Zhang, J.; Mukherjee, B.: "Network Performance of Optical Burst/Packet Switching: The Impact of Dimensioning, Routing and Contention Resolution," ITG-Fachtagung Photonische Netze, (Leipzig, Germany), 2005.
- [3] Ince, A.N.; Bragg, A.: "Recent Advances in Modeling and Simulation Tools for Communication Networks and Services," Springer, 2007.
- [4] Heidemann, J.; Mills, K.; Kumar, S.: "Expanding confidence in network simulations," IEEE Network, vol. 15, no. 5, pp. 58-63, September-October 2001.
- [5] K. Zhu and B. Mukherjee. On-line approaches for provisioning connections of different bandwidth granularities in WDM mesh networks. In Proceedings of the IEEE/OSA Optical Fiber Communication Conference (OFC), (Anaheim, CA, USA), March 2002.
- [6] Hülsermann, R; Jäger, M.; Gladisch, A.: "Improving the Efficiency of Dynamic Service Provisioning in Transport Networks with Scheduled Services," ITG-Fachtagung Photonische Netze, (Leipzig, Germany,) May 2-3, 2005.
- [7] Banerjee, A.; Wu-chun Feng; Ghosal, D.; Mukherjee, B.: "Algorithms for Integrated Routing and Scheduling for Aggregating Data from Distributed Resources on a Lambda Grid," IEEE Transactions on Parallel and Distributed Systems, Volume 19, Issue 1, Jan. 2008 Page(s):24 – 34.
- [8] Schupke, D.A.; Prinz, R.G., "Capacity Efficiency and Restorability of Path Protection and Rerouting in WDM Networks Subject to Dual Failures," Photonic Network Communications, Kluwer Academic Publishers, vol. 8, no. 2, pp. 191-207, September 2004.