TCP/IPoDWDM: A Case for Adaptive Congestion Control in Optical Networks under Physical Layer Impairments

Pedro P. P. Filho[1,2], Arnaldo P. Togneri[1,2], Moises R. N. Ribeiro[1], Marcelo E. V. Segatto[1], Everson Scherrer Borges[1,2], Magnos Martinello[1]

[1]Universidade Federal do Espirito Santo (UFES)
Av. Fernando Ferrari, 514 – 29.075-910 – Vitória – ES – Brazil

[2]Instituto Federal do Espirito Santo (Ifes), ES – Brazil

{pedro.pecolo,arnaldo.togner, everson}@ifes.edu.br
{moises.ribeiro,marcelo.segatto, magnos.martinello}@ufes.br

Abstract. *The IPoDWDM-driven network consolidation removes the demarcation between packet and optical layers, clearly increasing management system complexity, but also creating new opportunities. We propose TCP/IPoDWDM as a means to advance this consolidation, taking into account how physical layer impairments affect throughput and fairness depending on the TCP congestion control algorithm in use. This study is the first to explore dynamic congestion control algorithm adaptation at the server or virtual machine level as a potential solution for loss-sensitive applications, such as Data-Intensive Science (DiS). By leveraging Optical Signal-to-Noise Ratio (OSNR) metrics provided by integrated control and management planes, this approach aims to optimize performance during the session establishment phase. We outline the modeling approach, experiments conducted, and directions for future research.*

Resumo. *A consolidação de redes orientadas por IPoDWDM remove a demarcação entre as camadas de pacotes e óptica, aumentando significativamente a complexidade do sistema de gerenciamento, mas também criando novas oportunidades. Propomos o TCP/IPoDWDM como um meio de avançar nessa consolidação, considerando como as deficiências na camada física afetam a taxa de transferência e a imparcialidade, dependendo do algoritmo de controle de congestionamento TCP utilizado. Este estudo é o primeiro a explorar a adaptação dinâmica do algoritmo de controle de congestionamento no nível do servidor ou da máquina virtual como uma solução potencial para aplicações sensíveis a perdas, como a Ciência Intensiva em Dados (DiS). Ao aproveitar as métricas da Relação Sinal-Ruído Óptico (OSNR), fornecidas por planos de controle e gerenciamento integrados, esta abordagem visa otimizar o desempenho durante a fase de estabelecimento da sessão. Descrevemos a abordagem de modelagem, os experimentos conduzidos e as direções para pesquisas futuras.*

1. Introduction

The impact of open networking in academia is well recognized for research validation. In addition, governments and businesses are increasingly adopting it for transparency, security, and adaptability. The growing industry interest in open architectures is evident in the development of 400 OpenZR/ZR+ devices [1]. These advancements result from years of high-speed electronic and photonic integration, enabling 400G coherent systems in compact pluggable form factors [Hand et al. 2024]. Such technologies facilitate the long-anticipated shift towards Internet Protocol over Dense Wavelength Division Multiplexing (IPoDWDM) [Kamalzadeh et al. 2024], aiming to eliminate optical transponders to reduce cost, power, and space requirements. Which directly integrates IP with DWDM technology, simplifying network architecture while improving scalability, efficiency, and reducing latency, making it suitable for backbone networks and data-intensive applications. However, challenges arise, such as managing optical layer parameters within IP-based control and handling increased cross-layer interactions. Unlike traditional architectures, IPoDWDM exposes IP packets to optical impairments, leading to additional packet loss beyond conventional congestion events, which impacts higher-layer protocols [Hacker et al. 2002].

To address these challenges, we propose TCP/IPoDWDM, an architecture that adapts congestion control based on optical network state, allowing applications like Data Intensive Science (DiS) to leverage Optical Signal-to-Noise Ratio (OSNR) metrics for dynamic transport optimization made at session establishment phase. By enabling congestion control algorithms to switch in response to optical impairments, TCP/IPoDWDM prevents unnecessary overreaction to non-congestion-related packet losses. To validate this approach, a proper multilayer network representation is crucial. Emulation offers a more realistic means of studying complex network dynamics compared to traditional simulations. This paper contributes by: i) outlining the TCP/IPoDWDM proposal, ii) proposing a preliminary IPoDWDM implementation plan on Mininet, and iii) evaluating basic scenarios to assess transport adaptability impact using throughput and fairness metrics to refine congestion control mechanisms.

Section I brings background and related works. Section II briefly describes the physical layer issues. Sections II and III presents the description of the TCP/IPoDWDM system and test scenario, respectively. Section IV describes the emulated topology. Next, Sections V and VI address the results, conclusion and proposals for future works.

2. Background and Related Work

The consolidation enabled by IPoDWDM introduces both challenges and opportunities. As shown in Fig. 1, this evolution replaces traditional components (e.g., SR Optics, transponders) with OpenZR/ZR+ pluggables, directly connecting routers/switches to OLTs, Fig. 2. A key challenge is managing optical-layer parameters via IP-based control [Hand et al. 2024], while DCO transceivers enable real-time monitoring of fiber quality. [2]

In IPoDWDM, IP packet integrity is directly affected by physical layer impairments. Packet loss may result not only from congestion but also from the optical layer's current state. TCP, the dominant transport protocol, cannot distinguish between these causes. As a result, it reduces throughput and increases flow completion time. Its congestion control adjusts rates based on loss and RTT, misinterpreting physical-layer losses as

[1]Promoted by the Optical Internetworking Forum (OIF).

[2]Designing routed optical networks: `https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2023/pdf/BRKSPG-2029.pdf`.

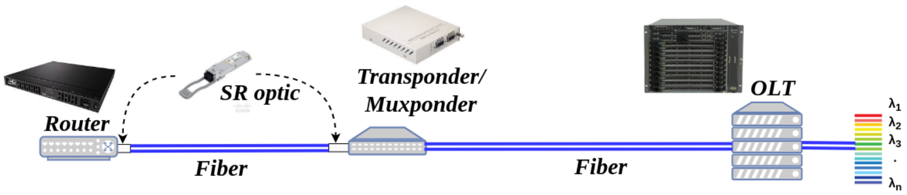


Figure 1. Traditional Optical Network

congestion. This limits performance under optical impairments [Hacker et al. 2002][3].

IPoDWDM pluggables support adaptive modulation (e.g., Table 1), adjusting bits/symbol based on OSNR, akin to WiFi's MCS [Peng et al. 2007]. However, unlike WiFi, optical-path adaptation must be centrally managed, not user-driven. The TCP/IPoDWDM proposal may decouple this dependency, enabling transport-layer optimizations per flow (at session establishment phase) by selecting appropriate congestion control mechanism for a given physical layer state.

3. TCP/IPoDWDM System: An Adaptive Congestion Control

We propose utilizing transport layer functionalities to mitigate cross-layer OSNR impacts for servers, virtual machines or individual flows, connected to the packet layer, as illustrated in Fig. 2.

A modern approach would be use AI-driven transport protocols, which could utilize control plane data from pluggables for better performance. But their deployment would require extensive validation, standardization, and system-wide kernel integration. We take here a more pragmatic approach. It involves dynamically adapting existing congestion control algorithms to real-time optical network conditions at session establishment phase [Kong et al. 2018]. Our initial goal is to match congestion control with OSNR feedback from pluggables.

This proof-of-principle study focuses on a single optical hop to assess potential benefits before full protocol development. As shown in Fig. 2, a local agent at the sender (e.g., virtual machine or container) can collect control/management data. We argue that congestion control algorithm may be adapted, from a list of well-established flavors, to improve TCP performance (e.g., throughput and fairness) coupling that decision to the physical layer conditions (e.g., OSNR). Taking consideration the described scenario about how características of physical layer can impact in behave of transport layer, and consequently the manner of how TCP can lead to these, this paper bring a propose to use the paramets of physical layer, ready by transceivers, and statistics get by tools like LACP (Link Aggregation Control Protocol), and use this information to switch to suitable TCP flavor, by a agent that can interact with source machine.For achieving this goal, however, one requires suitable modeling and emulation tools tailored for TCP/IPoDWDM integrated representation.

3.1. Physical Layer Impairments and Packet Loss Modeling

Packet losses in an optical link are directly dependent on the Bit Error Rate (BER), which reflects the probability of incorrect bit detection in a digital signal [Lathi and Ding 2010]. Assuming Gaussian Noise (AWGN) with zero mean and unit variance, the packet loss

[3] Although QUIC is relevant, it remains out of the scope of this analysis.

probability is related to BER as follows [Balam and Gibson 2006]: $Pr = 1-(1-BER)^L$, where Pr represents packet loss probability and $L = R.N$ the number of bits per packet, with N as symbols per packet and R as bits per symbol ratio.

Fiber optic link performance, particularly in long-distance transmission [Hartling et al. 2021], is commonly evaluated using Optical Signal-to-Noise Ratio (OSNR). [4] These metrics are combined into the generalized OSNR (gOSNR) [Hartling et al. 2021].

Studies such as [Yang et al. 2023] model BER for QPSK and M-QAM, highlighting key optical impairments that affect BER. [5]

3.2. Integration of Packet and Optical Layers

The Mininet Optical is an extension package for Mininet that adds transmission models for optical media. It enables the emulation of physical layer limitations and behaviors in optical networks, including the data plane in both optical and packet networks, as well as discrete components such as transceivers, amplifiers, fiber links, and ROADMs [Lantz et al. 2020].

The operating characteristics of OpenZR/ZR+, as well as the statistics collected by it, are compatible with the execution and modeling characteristics of Mininet-optical, as illustrated in Table 1 as in this study we operate a single wavelength channel.

Table 1. Optical parameters

Parameters	OpenZR+ (JCO400-QDD-ZR)	Mininet optical
Frequency	C-Band , 196.1 To 191.3 THz	C-Band , 196.1 To 191.3 THz
Modulation	16QAM, 8QAM, QPSK	64QAM, 16QAM
Channel spacing	75 GHz or 100 GHz	50 GHz
Transmitted Power	4 dBm To -6 dBm	-

4. Designing simple emulated test scenarios for explainable results

A Dumbbell topology was used, with four hosts, access and core switches, and a central router (Fig. 2). The setup demonstrates how physical layer impairments like ASE and NLI noise impact TCP performance, especially in overlapping flows where TCP RENO shows reduced fairness. To study the influence of delay, RTTs of 2, 20, 100, and 150 ms were selected based on [Chen et al. 2020]. No background traffic was added to focus on TCP behavior under bit-error-induced losses. We selected RENO, BIC, CUBIC, and BBR for evaluation: RENO, BIC, and CUBIC are loss-based, suitable for analyzing traditional responses to non-congestion-related loss, with BIC and CUBIC tuned for high-speed networks. BBR, combining loss and delay metrics, offers a modern contrast, enabling bandwidth estimation and lower latency. This selection provides a broad perspective on TCP performance in lossy scenarios.

The network emulation was conducted on a server running 64-bit Linux Ubuntu 22.03 (*kernel 6.2.0-39-generic*), with 48 GB RAM, a 12-core processor, and Mininet

[4]OSNR includes ASE and NLI noise, where $OSNR_{ASE}$ represents ASE noise interference within a 0.1 nm channel band, and $OSNR_{NLI}$ accounts for NLI noise interference.

[5]Instead of explicitly modeling ASE noise, NLI, and XPM, this study uses Mininet-Optical and GNPy for transmission modeling in DWDM-based optical systems.

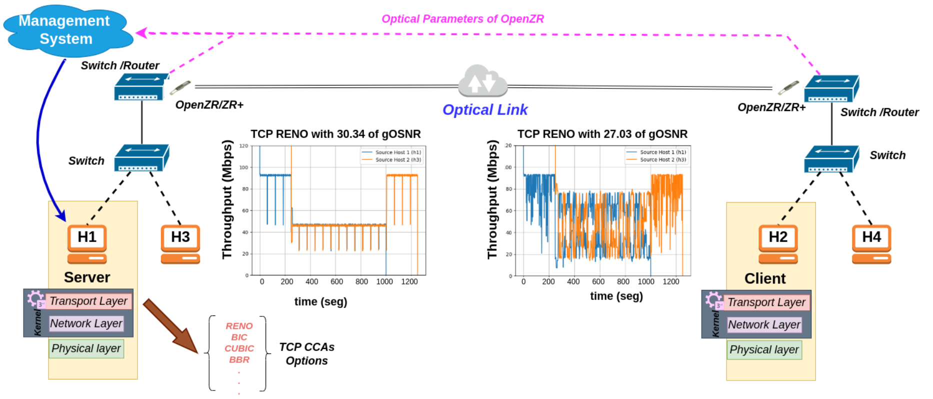


Figure 2. Topology and representative outcomes two competing (and partially overlapping) flows (H1-H2 blue and H3-H4 orange) using TCP RENO within two gOSNR scenarios and RTT 20 ms.

2.3.1b4. Besides TCP RENO inset Fig 2. During the experiments, average CPU usage ranged from 5 to 10%. Although multiple runs for ensured consistency, first we present a member-function of this stochastic process, instead of averaged data, to highlight specific TCP congestion control behaviors. This qualitative approach is preferred at this stage to better visualize the different congestion control dynamics under noise conditions [6].

5. Loss-Tolerant Congestion Control Performance for TCP/IPoDWDM

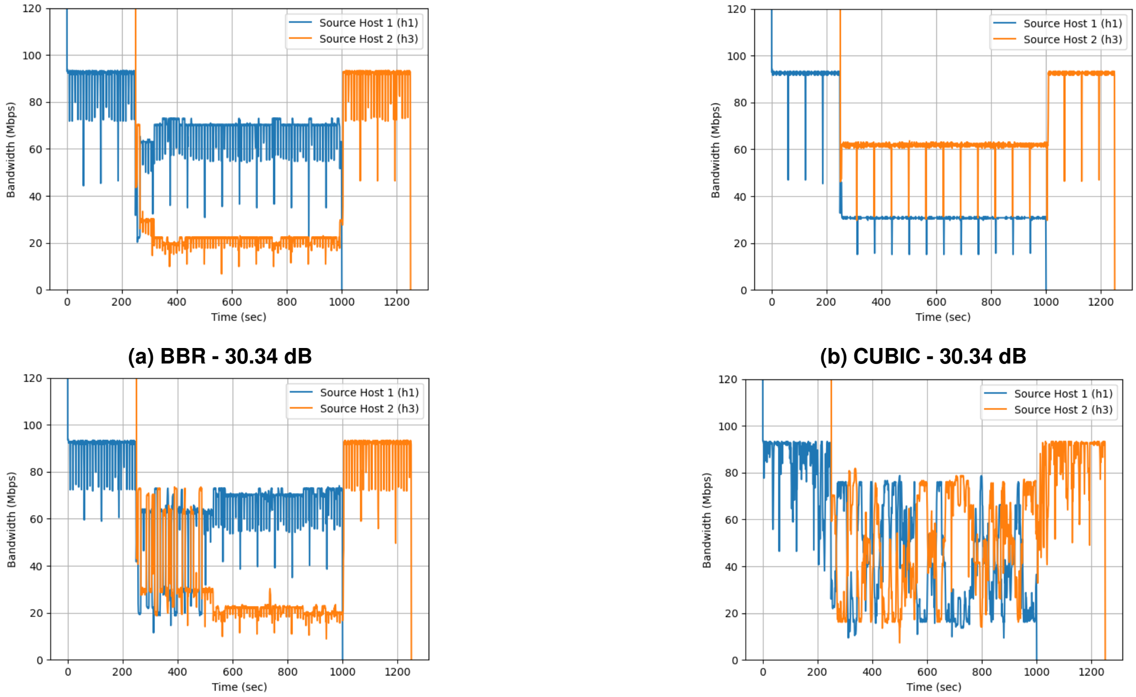
TCP RENO operating under reduced gOSNR will imply in continuous *cwnd* adjustments. This will be impacting, as a result, the instantaneous flow throughput. It is expected that CUBIC and BBR better handle such poor gOSNR due its less conservative *cwnd* management.

Fig. 3 presents CUBIC and BBR, as example, under different levels of gOSNR. Here we can observe how medium impairment can impact the throughput and how the flow interact with each other. With a decrease of 2.69 dB, it is possible to notice the degradation of both TCP BBR and CUBIC, although they still achieve a higher throughput than others TCPs for the same conditions.

It is clear that excessive random losses, induced by poor gOSRN (i.e., 27.65 dB), may cause instabilities in the fair-share mechanism of TCP congestion control algorithms. This is reflected in short-term throughput fluctuations. Therefore, a more comprehensive analysis should also consider Jain's fairness index. In this case, only the period during which TCP flows coexisted was considered; thus, the calculations are based on 10-second intervals. Moreover, for more conclusive outcomes, the mean and standard deviation of throughput, considering the stochastic nature of competing flows, are also computed within these overlapping 10-second intervals. Finally, the impact of RTT on the various congestion control algorithms needs to be considered in our investigation.

First, a baseline is provided in Fig. 4a where no optical layer was used. In Fig. 4b, the behavior of the TCPs for a condition of excellent channel quality (gOSNR $\geq$ 30 dB) is illustrated. A completely different scenario is observed in Fig.4c, despite only a 2.69 dB reduction in gOSNR. The change in algorithm behavior is notable, with an increase in Jain's Fairness Index [Sediq et al. 2013] for certain RTT values, although accompanied

[6] https://github.com/piccoli87/TCP-IPoDWDM.git



(a) BBR - 30.34 dB **(b) CUBIC - 30.34 dB**

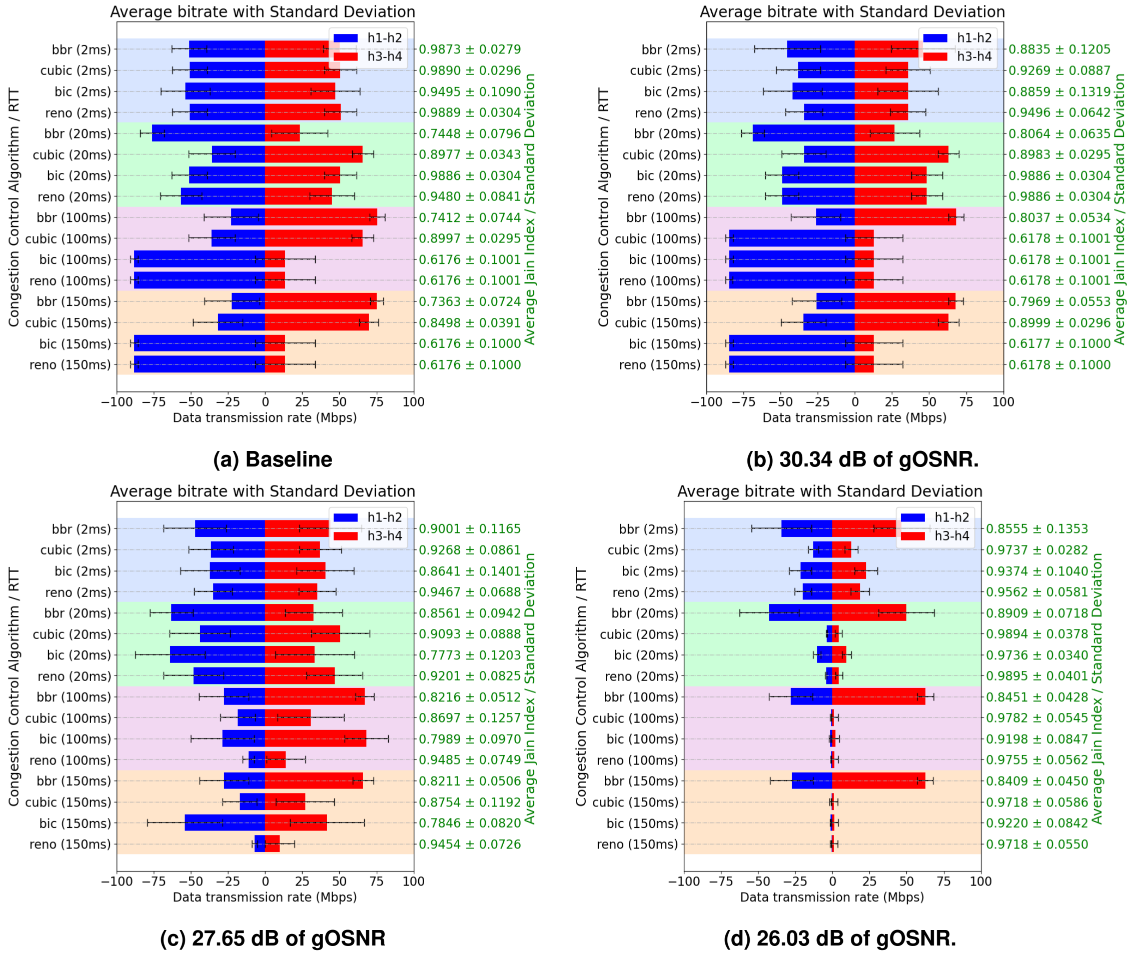
(c) BBR - 27.65 dB **(d) CUBIC - 27.65 dB**

Figure 3. Congestion behavior in optical network, 20 ms of RTT.

by the expected reduction in average bit rate, as seen in Figs.4b and 4c. This behavior leads to highly intermittent traffic for some flows. In Fig.4d, where gOSNR is around 26 dB, IPoDWDM performs poorly. However, the resilient behavior of BBR is noteworthy, followed by BIC under certain RTT conditions, while other TCP variants collapse under excessive packet loss in the optical layer. Notably, BBR and BIC exhibit a trend toward better fairness in TCP/IPoDWDM as RTT values increase, a behavior not consistently observed in other TCP variants.

This behavior arises from loss-based congestion algorithms, which gradually increase the transmission window until losses occur, then decrease it to stabilize without packet loss. By attempting to utilize all available link capacity, these algorithms lead to higher queue occupancy, increased delay, and packet loss due to overflow. Algorithms such as RENO, BIC and CUBIC base the adjustment of the transmission window on packet loss, with their behavior differing according to the form and reaction time to packet loss. The highlight of BBR is because, unlike other congestion control algorithms applied to TCP, it does not act as loss-bassed, but rather, what it calls of Bandwidth Bottleneck (*BtlBw*) and round trip proposals operating time (*RTprop*), [Yang et al. 2023].

This advantage of BBR is associated with its characteristic high retransmission rate. However, to minimize retransmissions while maintaining flexibility in media quality requirements, the use of FEC tools implemented in transceivers, such as OpenZR/ZR+, can overcome this limitation. TCP congestion control algorithms like BIC and CUBIC gradually increase the transmission rate until packet loss occurs. They then adjust the congestion window between the previous loss-free window (minimum) and the loss-inducing window (maximum) [Ha et al. 2008]. RENO, in contrast, reduces the congestion window (*cwnd*) to 1 upon loss detection. BBR follows a different approach, adjusting *cwnd* based on Bandwidth Bottleneck (*BtlBw*) and round-trip propagation time (*RTprop*), maintaining the transmission rate even under transient congestion [Yang et al. 2023].



(a) Baseline **(b) 30.34 dB of gOSNR.** **(c) 27.65 dB of gOSNR** **(d) 26.03 dB of gOSNR.**

Figure 4. Statistics of Congestion Control behavior: Throughput and Jain Index.

BBR's characteristics may explain its superior performance in low gOSNR scenarios. However, under high gOSNR, competing flows struggle to achieve fair-share throughput. Random losses from poor gOSNR affect both congestion window self-synchronization and fairness convergence. In conclusion, our results suggest that using gOSNR and RTT as a metric can help select the best TCP variant, with CUBIC excelling in fairness and BBR in throughput.

6. Conclusion

We proposed TCP/IPoDWDM as a step forward in the trend toward packet and optical network consolidation, considering that throughput and fairness may be affected differently by physical layer impairments depending on the TCP congestion control algorithm in use. This paper investigated whether dynamically adapting the congestion control algorithm could be a promising solution for loss-sensitive applications, such as DiS, especially now that metrics like gOSNR can be reported by integrated control and management planes. Modeling and explainable experiments preliminarily identified CUBIC and BBR as strong candidates for implementing our prototypes in the next phase of research toward TCP/IPoDWDM. Future work may also explore the impact of multiple channels on the same fiber and how nonlinear effects, such as XPM (Cross-Phase Modulation), and mixed TCP variants competing for bandwidth in multi-hop scenarios affect overall performance.

References

Balam, J. and Gibson, J. D. (2006). Path diversity and multiple descriptions with rate dependent packet losses. In *Proceedings of the Information Theory and Applications Workshop*.

Chen, X., Kim, H., Aman, J. M., Chang, W., Lee, M., and Rexford, J. (2020). Measuring tcp round-trip time in the data plane. In *Proceedings of the Workshop on Secure Programmable Network Infrastructure*, SPIN '20, page 35–41, New York, NY, USA. Association for Computing Machinery.

Ha, S., Rhee, I., and Xu, L. (2008). Cubic: A new tcp-friendly high-speed tcp variant. *SIGOPS Oper. Syst. Rev.*, 42(5):64–74.

Hacker, T. J., Athey, B. D., and Noble, B. (2002). The end-to-end performance effects of parallel tcp sockets on a lossy wide-area network. In *Proceedings 16th International Parallel and Distributed Processing Symposium*, pages 10–pp. IEEE.

Hand, S. J., Welch, D. F., Bhaskara, V., Naik, D., Wikman, A., Silvola, M., Su, H., Monteiro, T., Marques, F., Santos, P., et al. (2024). A new operational paradigm for ipodwdm networks. *IEEE Access*.

Hartling, E. R., Pilipetskii, A., Evans, D., Mateo, E., Salsi, M., Pecci, P., and Mehta, P. (2021). Design, acceptance and capacity of subsea open cables. *Journal of Lightwave Technology*, 39(3):742–756.

Kamalzadeh, M., Zhang, T., Fumagalli, A., and Manzotti, R. (2024). An ipodwdm implementation leveraging openroadm standards for enhanced service orchestration. In *2024 24th International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE.

Kong, Y., Zang, H., and Ma, X. (2018). Improving tcp congestion control with machine intelligence. In *Proceedings of the 2018 Workshop on Network Meets AI & ML*, pages 60–66.

Lantz, B., Díaz-Montiel, A. A., Yu, J., Rios, C., Ruffini, M., and Kilper, D. (2020). Demonstration of software-defined packet-optical network emulation with mininet-optical and onos. In *Optical Fiber Communication Conference (OFC) 2020*, page M3Z.9. Optica Publishing Group.

Lathi, B. P. and Ding, Z. (2010). *Modern digital and analog communication systems*. Oxford University Press, Inc.

Peng, F., Zhang, J., and Ryan, W. E. (2007). Adaptive modulation and coding for ieee 802.11 n. In *2007 IEEE Wireless Communications and Networking Conference*, pages 656–661. IEEE.

Sediq, A. B., Gohary, R. H., Schoenen, R., and Yanikomeroglu, H. (2013). Optimal tradeoff between sum-rate efficiency and jain's fairness index in resource allocation. *IEEE Transactions on Wireless Communications*, 12(7):3496–3509.

Yang, S., Tang, Y., Pan, W., Wang, H., Rong, D., and Zhang, Z. (2023). Optimization of bbr congestion control algorithm based on pacing gain model. *Sensors*, 23(9):4431.