A Survey of Multipath Transport Mechanism in Data Center Networks

Ruiwen Ji(✉), Lejun Ji, Keyang Gu, Junyi Wu, and Gang Lei

School of Software, Jiangxi Normal University, Nanchang 330022, China
jiruiwen@jxnu.edu.cn

Abstract. With the profound revolution in the data center industry and the increasing size of data center networks (DCNs), researchers have been focusing on achieving efficient and fast transport of traffic within the data center. In addition, the number of devices equipped with multiple interfaces is increasing, which allows multiple paths to be used simultaneously. Multipath transmission mechanisms can significantly improve the performance of data transmission in DCNs. However, some issues remain when MPTCP is deployed in DCNs. We have investigated multipath transport mechanisms in data center networks. First, the Multipath TCP (MPTCP) protocol is briefly described. Then, a study of routing algorithms, congestion control, and energy-saving techniques for applying MPTCP in DCNs are summarized separately. Finally, we propose the future direction of the multipath transmission mechanism in DCNs.

Keywords: data center networks · multipath TCP · software defined network

1 Introduction

The fourth industrial revolution driven by digital technology as the core is bringing profound changes to human production and life. As a physical base for carrying various digital technology applications, data centers are gradually highlighting their industrial empowerment value [1]. Major countries around the world are actively guiding the development of the data center industry. The data center market is growing in size and traffic convergence within data center networks (DCNs) is also increasing. If the DCNs experience network congestion, there may occur problems such as communication delays, long query times, and degraded service quality, causing a series of significant losses [2,3]. To avoid traffic surges that strain data centers and affect the quality of service, achieving efficient and fast transmission of traffic within DCNs has become a hot research topic today.

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61962026, and by the Postgraduate Innovation Fund of Jiangxi Provincial Department of Education under Grant YC2021-S258.

Y. Cao and X. Shao (Eds.): MONAMI 2022, LNICST 474, pp. 373–385, 2023.
https://doi.org/10.1007/978-3-031-32443-7_27

As different kinds of network access technologies flourish, the next-generation network will be a heterogeneous network with multiple access networks coexisting in wired, wireless, and cellular networks [4]. With the accelerated deployment of IPv6, multiple addresses of hosts will become more popular. The servers in DCNs are basically multi-homed hosts, and there are usually multiple available paths between two hosts. However, traditional TCP can only select one path at random, which wastes a lot of network resources. Designed as an extension to TCP, Multipath TCP (MPTCP) [5] takes advantage of multiple subflows on different paths to enable high throughput and dynamic traffic migration. There are many solutions based on MPTCP [6–9] that have shown particular advantages for viable deployment, low latency and high throughput in DCNs. In brief, MPTCP can achieve load balancing by creating multiple subflows to distribute data to multiple paths within the network [10]. Figure 1 shows an MPTCP communication scenario under DCNs where two multi-homed hosts (Host A and Host B) can establish connections over three paths, meaning that multiple paths can be used to exchange data between the hosts simultaneously. This kind of multipath solution increases throughput and builds higher resistance to network failures [11].

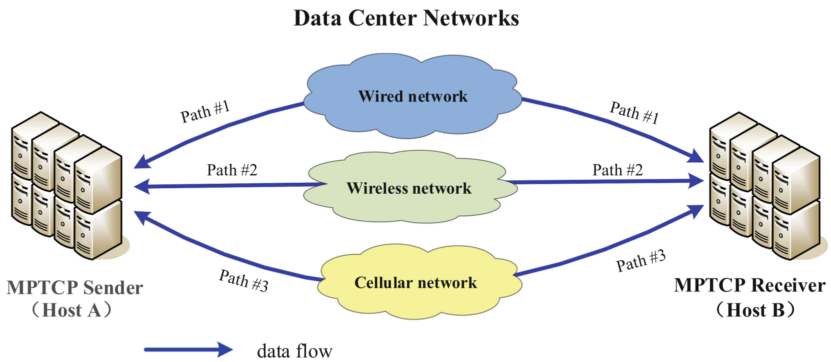

Fig. 1. MPTCP communication scenarios in DCNs.

However, some problems still exist when MPTCP is deployed in DCNs. For instance, due to the multi-versus-one model and the abundance of short pulses in DCNs, all traffics are prone to collision on shared links and on the same output port. Because the buffers of ToRs are usually too small to handle bursts, MPTCP incast problems inevitably occur during transmission [12], causing bottleneck collapse and dramatic throughput degradation. In addition, it is shown that in modern DCNs, MPTCP is effective over long traffic but not over short traffic [6]. Massive timeout retransmissions usually cause short traffic to arrive late. Which significantly reduces the transmission efficiency of a large number of delay-sensitive short flows in DCNs. These network topologies with multipath characteristics can generate large resource wastage when the network load is low.

Since the energy consumption of current DCNs devices and components is less affected by load variation, they consume essentially similar energy at low loads as at full loads [13], and the more underutilized devices can significantly pull down the efficiency of the network. The huge energy consumption of DCNs leads to high electricity costs, constrains the expansion of DCNs, and exacerbates the damage and impact on the environment [14]. Therefore, the performance and energy consumption issues of multipath transmission in DCNs have aroused widespread concern in academia and industry.

This paper surveys the research on multipath transport mechanisms in DCNs, and proposes future development directions and challenges. The survey is organized as follows. Section 2 introduces the MPTCP overview. Section 3 reviews MPTCP technologies in DCNs. Section 4 discusses the directions for future research. And we summarize in Sect. 5.

2 MPTCP Overview

With the continuous development of communication technology, modern communication terminals generally have multiple network interfaces. However, traditional TCP can only utilize a single interface for communication during one data transmission. To fully utilize the network resources among multi-homed hosts, in 2013, the International Internet Engineering Task Force (IETF) proposed the MPTCP [4], and standardized the protocol stack design, architecture, congestion control, application interfaces, and usage scenarios, etc. The MPTCP enables multi-homed terminals to access multiple networks simultaneously and fit multiple link bandwidths to achieve multi-path data transmission, so as to increase data transfer rates and maximize network resource utilization.

As shown in Fig. 2, the MPTCP layer is located below the application layer and above the TCP layer in the protocol stack [15], providing a standard TCP interface for the application layer to hide multipaths while multiple TCP subflows need to be managed. With functions such as path management [16], packet scheduling [17] and congestion control [18], MPTCP is currently the most researched and widely used transport layer multipath transport technology. MPTCP has attracted a lot of attention in the DCN field due to the following advantages.

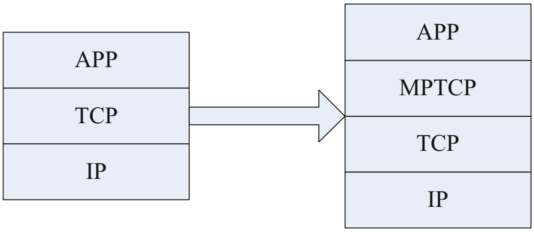

Fig. 2. MPTCP protocol stack.

The main advantages of the MPTCP protocol are as follows.

- **Improve throughput.** Traditional TCP only supports single-path data transmission between communicating pairs. MPTCP can take full advantage of multiple hosts equipped with multiple network interfaces to create multiple TCP subflows to transmit data in parallel between communicating pairs, fully aggregating the bandwidth on different physical links and improving throughput.
- **Enhance robustness.** In the event of poor link communication quality or disconnection, traditional TCP must disconnect the current connection and establish a new TCP connection on another link to re-transmit the data. For MPTCP, when a subflow fails, MPTCP management can quickly and seamlessly switch services to other available subflows, maintaining continuous data transmission and improving resilience to dynamic network environments.
- **Congestion Balance.** MPTCP assigns data to different subflows for transmission based on the subflow status and transfers the more congested subflows to the idle subflows through the congestion control mechanism to achieve load balancing of the entire network and improve the resource utilization of the network.
- **Traditional TCP compatibility.** MPTCP is an extension of traditional TCP. When the communication counterpart supports the MPTCP, it is preferred to use the MPTCP for data transmission to take advantage of multipath; when the communication counterpart does not support the MPTCP, it can also directly degrade to the traditional TCP to continue the single-path transmission.
- **Other Compatibility.** The MPTCP addresses compatibility issues including with the application layer, network middleware, and other users in the network, and ensures the security of data transmission in the face of attacks [19].

3 Status of Research on Multipath Transport Mechanisms in DCNs

3.1 Data Center Networks

The data center acts as an infrastructure service provisioning role to deliver, compute, and store data information. The traditional architecture of DCNs is organized into three layers [20], as shown in Fig. 3, mainly consisting of a large number of Layer 2 access devices and a small number of Layer 3 devices with a tree-like layered structure. Core Layer is the high-speed switching backbone that enables optimized data transmission between backbone networks through high-speed forwarding. Aggregation Layer is in the middle of Core Layer and Access Layer, and is responsible for processing and aggregating all communication data flows in the uplink of Core Layer and downlink of Access Layer, as well as providing firewall and intrusion detection services. Access Layer, also known

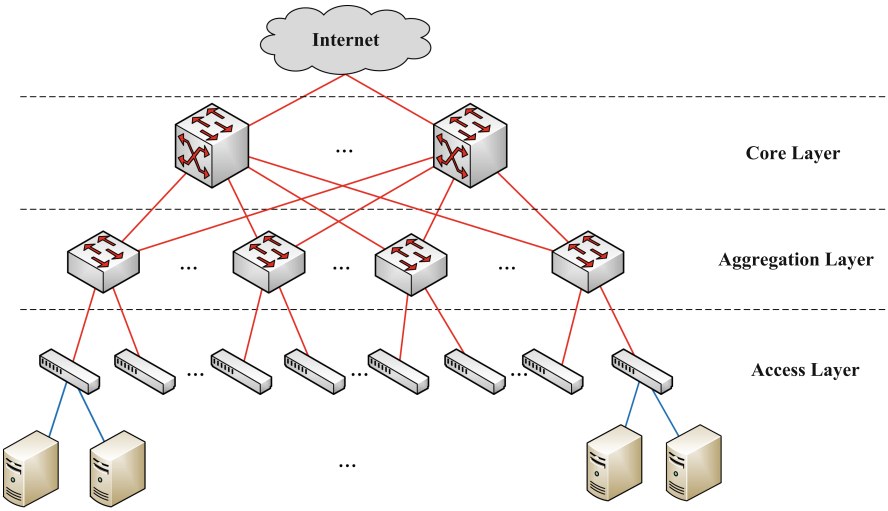

Fig. 3. The topology of traditional DCNs.

as Edge Layer, is mainly responsible for the access of network devices such as servers.

In recent years, with the rapid growth of data center traffic and the application requirements of virtualization technology, DCNs bandwidth and performance are facing great challenges. The traditional network architecture is no longer suitable for the development of the new generation of data centers, so researchers have designed a variety of network architectures, such as fat-tree [21] and VL2 [22] with switches as the core, and Dcell [23] with servers as the core.

In DCNs, the flows with high bandwidth demand are called long flows (also called elephant flows), which are mainly generated by data backups, and the flows with less bandwidth demand but more sensitive to latency are called short flows (also called rat flows), which are mainly generated by online searches [24]. These two types of traffic in the data center are distributed in a heavy-tailed manner. 90% of the volume of the network is rat streams, which accounts for less than 10% of the whole network's traffic. The other less than 10% of the number of elephant streams accounts for more than 80% of the total network traffic.

3.2 Routing Algorithms for Multipath Transmission in DCNs

Due to the popularity of multipath DCNs topologies, researchers have proposed many routing algorithms to solve the load balancing problem of DCNs. The equal-cost multipath routing (ECMP) algorithm [25] is the main routing algorithm used in DCNs to solve load balancing problems, which first numbers multiple equivalent paths, then hashes the packet header fields, and finally maps the data flow to the corresponding path. However, the ECMP algorithm may map multiple long flows onto the same link, resulting in link congestion and packet loss. The ECMP algorithm performs well when the network load is light, but as the network load increases and the network fluctuations increase, the load on the network becomes unbalanced. ECMP has two obvious disadvantages, (i) ECMP does not have a congestion awareness mechanism, and continues to assign forwarding tasks to already congested links, resulting in increased congestion; (ii) ECMP does not distinguish between elephant flows and rat flows, and this indiscriminate forwarding strategy is obviously no longer applicable to today's DCNs. RPS [26] maps every packet randomly to all available paths to optimise the resources of the multipath, but this approach can easily lead to packet reordering. To prevent packet reranking, CAPS [27] handles short flows and propagates them to all paths. For avoiding drastic reordering, Hermes [28] only reorders packets to good paths in a timely and careful manner when it is advantageous to do so. The DLB [29] algorithm is a dynamic load balancing algorithm for fat-tree topology networks, DLB adopts the idea of greedy strategy, starting from the sending end and selecting the other node of the link with the largest remaining bandwidth as the next hop until it reaches the highest-level switch that the flow needs to reach. However, the path obtained by the DLB algorithm is only locally optimal, which may cause congestion of some locally optimal paths. He, K. *et al.* [30,31] improved network elasticity by making subflows not pass overlapping links in DCNs. However, the routing scheme only considers the current single connection and completely ignores the other connections in the network, and cannot do anything about the collisions between the subflows of different connections.

MPTCP transmits data in subflows, making full use of multipath resources. A number of works have emerged to improve the performance of MPTCP in DCNs. Raiciu, C. *et al.* [32] demonstrate that multipath transmission control protocols, while fully utilizing bandwidth and increasing throughput, increase finish times for short flows. Cao, Y. *et al.* [33] Achieve high data transmission performance by dynamically constructing disjoint paths for each MPTCP flow to take advantage of the path diversity of DCNs. However, MPTCP seeks to realise enough link usage by depleting buffers, resulting in sizeable queuing delays and packet loss. This affects the performance of the short, latency-sensitive traffic. Therefore, XMP [34] controls switch buffer occupancy using the ECN mechanism to lessen the influence of long flows against short flows. MMPTCP [6] differs long and short flows based on the amount of bytes, as well as applying different transmission policies.

As a network innovation architecture, Software Defined Network (SDN) realizes the decoupling of control plane and data plane under the traditional network architecture, as shown in Fig. 4. In theory, SDN can explore, allocate, and manage network resources for each MPTCP connection in a fair, efficient, and fast manner. In the DCNs based on SDN, Duan, J. *et al.* [35] designed a responsive MPTCP path management system, which consists of two modules: one is the centralized control module, which is responsible for the intelligent computing function of subflow routes; the other is the monitoring module, which proactively tunes the number of subflows depending on the network requirements. Hussein, A. *et al.* [36] designed an enhanced MPTCP architecture using SDN technology to achieve the target of maximising DCNs' throughput. However, collisions between different subflows of the same connection will affect transmission efficiency. Zannettou, S. *et al.* [37] designed a disjoint allocation strategy for MPTCP subflows. By parsing the MPTCP packet header information, the SDN controller assigns different subflows of the same connection to non-overlapping paths, thus making full use of network bandwidth. However, these methods ignore the quality difference between paths, which leads to serious packet disorder and a sharp decline in transmission performance. Kukreja, N. *et al.* [38] proposed the design of an MPTCP path manager based on SDN. With an intensive analysis of different scheduling algorithms, they discussed the impact of delay differences on the general performance of MPTCP and effectively mitigated the performance degradation caused by path quality differences.

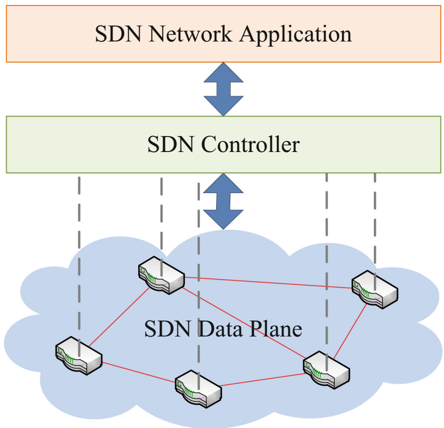

Fig. 4. SDN Architecture.

3.3 Congestion Control for Multipath Transmission in DCNs

More and more studies have shown that multipath transport technology would become the critical technology for DCNs. However, MPTCP has a problem with incast crashes, and its throughput drops drastically when the receiver makes data requests on multiple servers at the same time. Researchers have proposed many congestion control mechanisms to prevent MPTCP's incremental crash problem and thus improve MPTCP's performance in DCNs.

Kheirkhah, M. *et al.* [6] analysed the obstacles to deploying MPTCP in DCNs. The authors proposed a multipath congestion control scheme that can adaptively switch between multi-run and single-run traffic. Ye, J. *et al.* [39] proposed an enhanced MPTCP protocol that can adjust the congestion control mechanism under different numbers of subflows. Kimura, B. *et al.* [40] proposed an algorithm to adjust the send operations of all subflows in a collocated way to accomplish diverse goals such as friendliness, improved throughput and congestion balancing. However, this method does not reduce the number of lost packets. QCBF [41] is a multi-homed fat-tree topology MPTCP transport control method based on ToR cluster buffer assignment, which effectively eliminates the MPTCP incremental problem. DCMPTCP [8] controls the creation of subflows to alleviate congestion in data center communication mode. MMPTCP [6] randomly disperses short flow packets to allow for high aggregated throughput to balance the traffic load across multiple paths. FUSO [42] modifies the retransmission mechanism of MPTCP subflows. In cases where the sender suspects that a packet is lost, it will rapidly retransmit the lost packet using other congestion windows with small loads, cutting down on the total completion time of the subflow. MPTCP_OPN [44] adapts the number of parallel subflows based on real-time network conditions and moves traffic away from crowded paths flexibly to alleviate high latency. EW-MPTCP [12] allows additional congestion control operations for each subflow, mitigating incast crashes by permitting several MPTCP subflows to compete equitably with a single TCP flow at a common bottleneck.

3.4 Energy-Saving Techniques for Multipath Transmission in DCNs

Data centers around the world consume a lot of electricity, generating huge carbon emissions. It is well known that the energy consumption of data communications relies on the CPU power of the host and the traffic finish time. The instantaneous CPU power of the host during data transfer increases significantly compared to the free state and is influenced by factors such as the sending rate and the number of network interfaces. The completion time of elephant streams falls with increasing throughput, while the completion time of rat streams relies strongly on the quality of the path in real time. Therefore, energy optimisation for data transmission must take all these factors into account.

The reduction in completion time for long flows and the energy consumption using MPTCP compared to single-path TCP more than offset the increase in host CPU power. Researchers have proposed a variety of data center network energy-saving algorithms. Zhao, J. *et al.* [45] devised an energy-efficient approach for data centers, MPTCP-D. It reduces energy consumption by minimising traffic fulfilment time and closes the congestion window for additional subflows, guaranteeing that only one subflow remains on the overlapping path. Raiciu, C. *et al.* [32] demonstrated that MPTCP can improve data center utilisation at different network sizes and topologies. Khalili, R. *et al.* [46] studied the problem of optimal resource assignment between vast numbers of synchronous MPTCP flows. Interference of different MPTCP subflows on a common link raises the cost of traffic management overhead and increases energy consumption. If packet loss occurs, current approaches still need to wait for a retransmission mechanism to make a decision [47], which is less efficient for short delay-sensitive traffic. Gupta,

M. *et al.* [48] proposed that during periods of low activity, changing routes aggregates traffic onto fewer routes, causing idle devices to sleep. Vasić, N.*et al.* [49] proposed EATE to reduce energy consumption by combining sleep, rate adaptation, and routing coordination. Lin, M. *et al.* [50] studied data center energy proportional management to match energy consumption to server workloads.

4 Future Directions and Challenges

With the deployment of IPv6, the scale of multiple hosts in DCNs will be further expanded. At present, the research on multipath transmission mechanisms in DCNs mainly focuses on routing algorithms, congestion control, and energy-saving technology, which shows the advantages of multipath transmission technology. However, future research will face many challenges.

– **Design a multipath transmission strategy for hybrid flows in DCNs.** Many existing transport schemes first distinguish between elephant and rat flows in DCNs, and then perform path selection and data scheduling, respectively. However, there is currently no precise method for traffic detection and differentiation, and if an error is detected, it will seriously affect the data transmission performance. Future research should combine transmission control strategies with different traffic patterns, comprehensively consider switch caching, real-time link congestion, etc., design hybrid streaming transmission strategies for DCNs, and make balanced use of network resources.
– **Design the deployment scheme of large-scale SDN in DCNs.** As a new network architecture [51], the core idea of SDN is numerical control separation. The controller can create flexible network forwarding policies based on complex and variable service scenarios, providing a new mindset for data center traffic load balancing. MPTCP synchronous traffic arrivals can easily cause switch buffer overflows. Future work should address MPTCP load issues on SDN controllers, for example, by using multiple controllers to alleviate performance bottlenecks.
– **Study the energy consumption model of multipath transmission in DCNs.** With the expansion of the data center scale and business, energy consumption is a direct reflection of the cost. To the research community, the power consumption of MPTCP for DCNs remains largely unknown. In the future, energy-saving schemes should be studied to realize that the energy consumption of a data center is proportional to the workload level under the active state, and maximize the load balance and energy-saving effect by means of game theory.
– **Study the robustness optimization scheme of multipath transmission in DCNs.** Traffic bursts can be fatal to data center applications and can lead to serious performance degradation. If a critical transmission path in DCNs encounters random failures or external network attacks, the transmission performance of the path deteriorates, affecting the overall performance of the multipath transmission. At present, there are few studies on the security

of multipath transmission in DCNs. Therefore, future research can combine dynamic system modeling theory with complex network analysis methods to establish DCNs multipath transmission system robustness analysis model.

5 Conclusion

The multipath transmission mechanism will improve the transmission performance of DCNs. In recent years, great progress has been made in improving MPTCP performance in DCNs. This paper presents a current survey of the multipath transmission mechanism in DCNs. In order to analyse this field, we discuss the study of routing algorithms, congestion control, and energy-saving techniques for applying MPTCP in DCNs. In addition, we also put forward the development trend of the multipath transmission mechanisms in DCNs and the key problems to be solved to help researchers carry out further research.

References

1. Zhang, T., Lei, Y., Zhang, Q., Zou, S., Huang, J., Li, F.: Fine-grained load balancing with traffic-aware rerouting in datacenter networks. J. Cloud Comput. **10**(37), 1–20 (2021)
2. Xu, C., Wang, K., Sun, Y., Guo, S., Zomaya, A.Y.: Redundancy avoidance for big data in data centers: a conventional neural network approach. IEEE Trans. Netw. Sci. Eng. **7**(1), 104–114 (2020)
3. Huang, J., Li, S., Han, R., Wang, J.: Receiver-driven fair congestion control for TCP outcast in data center networks. J. Netw. Comput. Appl. **131**, 75–88 (2019)
4. Cao, Y., Ji, R., Ji, L., Lei, G., Wang, H., Shao, X.: l^2-MPTCP: a learning-driven latency-aware multipath transport scheme for industrial internet applications. IEEE Trans. Industr. Inf. (2022). https://doi.org/10.1109/TII.2022.3151093
5. Ford, A., Raiciu, C., Handley, M., Bonaventure, O.: TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6824, IETF (2013)
6. Kheirkhah, M., Wakeman, I., Parisis, G.: Multipath transport and packet spraying for efficient data delivery in data centers. Comput. Netw. **162**, 1–15 (2019)
7. Kheirkhah, M., Lee, M.: A solution to MPTCP's inefficiencies under the incast problem for data center networks. Comput. Commun. **161**, 238–247 (2020)
8. Dong, E., Fu, X., Xu, M., Yang, Y.: DCMPTCP: host-based load balancing for datacenters. In: 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), pp. 622–633. IEEE, Piscataway (2018)
9. Duan, J., Wang, Z., Wu, C.: Responsive multipath TCP in SDN-based datacenters. In: 2015 IEEE International Conference on Communications (ICC), pp. 5296–5301. IEEE, Piscataway (2015)
10. Mahmud, I., Lubna, T., Song, Y.-J., Cho, Y.-Z.: Coupled multipath BBR (C-MPBBR): a efficient congestion control algorithm for multipath TCP. IEEE Access **8**, 165497–165511 (2020)
11. Cao, Y., Ji, R., Huang, X., et al.: Empirical mode decomposition-empowered network traffic anomaly detection for secure multipath TCP communications. Mob. Netw. Appl. **27**, 1–10 (2022)
12. Li, M., Lukyanenko, A., Tarkoma, S., Ylä-Jääski, A.: MPTCP incast in data center networks. China Commun. **11**(4), 25–37 (2014)

13. Heller, B., Seetharaman, S., Mahadevan, P., et al.: ElasticTree: saving energy in data center networks. In: the 2010 USENIX Conference on Networked Systems Design & Implementation, pp. 249–264. USENIX Association, USA (2010)
14. Bhattacharya, T., Qin, X.: Modeling energy efficiency of future green data centers. In: 2020 11th International Green and Sustainable Computing Workshops (IGSC), pp. 1–3 (2020)
15. Ford, A., Raiciu, C., Handley, M., Barre, S., Iyengar, J.: Architectural guidelines for multipath TCP development. RFC 6182 (2011)
16. Chung, J., Han, D., Kim J., Kim, C.: Machine learning based path management for mobile devices over MPTCP. In: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 206–209. IEEE, Piscataway (2017)
17. Hwang, J., Yoo, J.: Packet scheduling for multipath TCP. In: 2015 Seventh International Conference on Ubiquitous and Future Networks, pp. 177–179. IEEE, Piscataway (2015)
18. Wei, W., Xue, K., Han, J., Wei, D., Hong, P.: Shared Bottleneck-based congestion control and packet scheduling for multipath TCP. IEEE/ACM Trans. Network. **28**(2), 653–666 (2020)
19. Nguyen, V.D., Ro, S.: Performance evaluation of MPTCP over shared bottleneck link. J. Korean Inst. Commun. Inf. Sci. **40**(1), 70–77 (2015)
20. Chen, K., Singla, A., Singh, A., et al.: OSA: an optical switching architecture for data center networks with unprecedented flexibility. In: the USENIX Symposium on Networked Systems Design and Implementation (NSDI), pp. 498–511. IEEE, Piscataway (2014)
21. Al-Fares, M., Loukissas, A., Vahdat, A.: A scalable, commodity data center network architecture. ACM SIGCOMM Comput. Commun. Rev. **38**(4), 63–74 (2008)
22. Greenberg, A., Hamilton, J.R., Jain, N., et al.: VL2: a scalable and flexible data center network. In: the Special Interest Group on Data Communication (SIGCOMM), pp. 51–62. Association for Computing Machinery, New York, NY, USA (2009)
23. Guo, C., Wu, H., Tan, K., Shi, L., Zhang, Y., Lu, S.: DCell: a scalable and fault-tolerant network structure for data centers. In: the ACM SIGCOMM 2008 Conference on Data Communication (SIGCOMM 2008), pp. 75–86. Association for Computing Machinery, New York, NY, USA (2008)
24. Benson, T., Akella, A., Maltz, D.A.: Network traffic characteristics of data centers in the wild. In: the 10th ACM SIGCOMM Conference on Internet Measurement, pp. 267–280. Association for Computing Machinery, New York, NY, USA (2010)
25. Hopps, C.: Analysis of an equal-cost multi-path algorithm, RFC2992 (2000). https://doi.org/10.17487/RFC2992. Accessed 01 Nov 2000
26. He, K., Rozner, E., Agarwal, K., Felter, W., Carter, J., Akella, A.: Presto: edge-based load balancing for fast datacenter networks. In: the Conference on Special Interest Group on Data Communication SIGCOMM 2015. Association for Computing Machinery, New York, NY, USA (2015)
27. Hu, J., Huang, J., Lv, W., Zhou, Y., Wang, J., He, T.: CAPS: coding-based adaptive packet spraying to reduce flow completion time in data center. IEEE/ACM Trans. Network. **27**(6), 2338–2353 (2019)
28. Zhang, H., Zhang, J., Bai, W., Chen, K., Chowdhury, M.: Resilient datacenter load balancing in the wild. In: The Conference of the ACM Special Interest Group on Data Communication SIGCOMM' 17. Association for Computing Machinery, pp. 253–266. NY, USA, New York (2017)

29. Yu, L., Deng, P.: Open flow based load balancing for fat-tree networks with multipath support. In: IEEE International Conference on Communications (ICC). IEEE, Piscataway (2013)
30. He, K., Rozner, E., Agarwal, K., Felter, W., Carter, J., Akella, A.: Presto: edge-based load balancing for fast datacenter networks. In: the Conference on Special Interest Group on Data Communication SIGCOMM' 15, pp. 75–86. Association for Computing Machinery, New York, NY, USA (2015)
31. Sandri, M., Silva, A., Rocha, L.A., Verdi, F.L.: On the benefits of using multipath TCP and openflow in shared bottlenecks. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 9–16. IEEE, Piscataway (2015)
32. Raiciu, C., Barre, S., Pluntke, C., et al.: Improving datacenter performance and robustness with multipath TCP. ACM SIGCOMM Comput. Commun. Rev. **41**(4), 266–277 (2011)
33. Cao, Y., Xu, M.: Dual-NAT: dynamic multipath flow scheduling for data center networks. In: 2013 21st IEEE International Conference on Network Protocols (ICNP), pp. 1–2. IEEE, Piscataway (2013)
34. Cao, Y., Xu, M., Fu, X., Dong, E.: Explicit multipath congestion control for data center networks. In: the Ninth ACM Conference on Emerging Networking Experiments and Technologies - CoNEXT' 13, pp. 73–84. Association for Computing Machinery, New York, NY, USA (2013)
35. Duan, J., Wang, Z., Wu, C.: Responsive multipath TCP in SDN-based datacenters. In: 2015 IEEE International Conference on Communications (ICC), pp. 5296–5301. IEEE, Piscataway (2015)
36. Hussein, A., Elhajj, I.H., Chehab, A., et al.: SDN for MPTCP: an enhanced architecture for large data transfers in datacenters. In: IEEE International Conference on Communications, pp. 1–7. IEEE, Piscataway (2017)
37. Zannettou, S., Sirivianos M., Papadopoulos, F.: Exploiting path diversity in datacenters using MPTCP-aware SDN. In: 2016 IEEE Symposium on Computers and Communication (ISCC), pp. 539–546. IEEE, Piscataway (2016)
38. Kukreja, N., Maier, G., Alvizu R., Pattavina, A.: SDN based automated testbed for evaluating multipath TCP. In: 2016 IEEE International Conference on Communications Workshops (ICC), pp. 718–723. IEEE, Piscataway (2016)
39. Ye, J., Feng, L., Xie, Z., Huang, J., Li, X.: Fine-grained congestion control for MultiPath TCP in data center networks. IEEE Access **7**, 31782–31790 (2019)
40. Kimura, B., Loureiro, A.: MPTCP linux kernel congestion controls. Technical report (2018)
41. Pang, S., Yao, J., Wang, X., Ding, T., Zhang, L.: Transmission control of MPTCP Incast based on buffer balance factor allocation in data center networks. IEEE Access **7**, 183428–183434 (2019)
42. Guo, C., Lu, Y., Yuan, M., et al.: Fast and cautious: leveraging multi-path diversity for transport loss recovery in data centers. In: USENIX Annual Technical Conference, pp. 29–42. USENIX Association, USA (2016)
43. Liu, S., Huang, J., Jiang, W., Wang, J.: Reducing traffic burstiness for MPTCP in data center networks. J. Netw. Comput. Appl. **192**, 1–12 (2021)
44. Huang, J., Li, W., Li, Q., Zhang, T., Dong, P., Wang, J.: Tuning high flow concurrency for MPTCP in data center networks. J. Cloud Comput. Adv. Syst. Appl. **9**(13), 1–15 (2020)
45. Zhao, J., Liu, J., Wang, H., Xu, C., Gong, W., Xu, C.: Measurement, analysis, and enhancement of multipath TCP energy efficiency for datacenters. IEEE/ACM Trans. Network. **28**(1), 57–70 (2020)

46. Khalili, R., Gast, N., Popovic, M., Upadhyay, U., Boudec, J.: MPTCP is not Pareto-optimal: performance issues and a possible solution. In: the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT 2012), pp. 1–12. Association for Computing Machinery, New York, NY, USA (2012)
47. Ferlin, S., Alay, Ö., Dreibholz, T., Hayes, D.A., Welzl, M.: Revisiting congestion control for multipath TCP with shared bottleneck detection. In: the 35th Annual IEEE International Conference on Computer Communications, pp. 1–9. IEEE, Piscataway (2016)
48. Gupta, M., Singh, S.: Greening of the internet. In: Conference on Applications. Technologies, Architectures, and Protocols for Computer Communications, pp. 19–26. Association for Computing Machinery, New York, NY, USA (2003)
49. Vasić, N., Kostić, D.: Energy-aware traffic engineering. In: the 1st International Conference on Energy-Efficient Computing and Networking (e-Energy 2010), pp. 169–178. Association for Computing Machinery, New York, NY, USA (2010)
50. Lin, M., Wierman, A., Andrew, L.L., Thereska, E.: Dynamic right-sizing for power-proportional data centers. IEEE/ACM Trans. Network. **21**(5), 1378–1391 (2013)
51. Hu, F., Hao, Q., Bao, K.: A survey on software-defined network and Openflow: from concept to implementation. Commun. Surv. Tutorials **16**(4), 2181–2206 (2014)