

References

- [1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A System for Large-Scale Machine Learning. In *OSDI*, 2016.
- [2] Amazon. Elastic Load Balancing. <https://aws.amazon.com/elasticloadbalancing/>.
- [3] Andrew Morgan. How MySQL is able to scale to 200 Million QPS - MySQL Cluster. <http://highscalability.com/blog/2015/5/18/how-mysql-is-able-to-scale-to-200-million-qps-mysql-cluster.html>.
- [4] Apache. Arrow. <https://arrow.apache.org/>.
- [5] Apache. Thrift. <https://thrift.apache.org/>.
- [6] M. T. Arashloo, A. Lavrov, M. Ghobadi, J. Rexford, D. Walker, and D. Wentzlaff. Enabling Programmable Transport Protocols in High-Speed NICs. In *NSDI*, 2020.
- [7] C. P. Authors. Cap'n Proto Cerealization Protocol. <https://capnproto.org/>.
- [8] T. Ball, N. Bjørner, A. Gember, S. Itzhaky, A. Karbyshev, M. Sagiv, M. Schapira, and A. Valadarsky. VeriCon: Towards Verifying Controller Programs in Software-Defined Networks. In *PLDI*, 2014.
- [9] A. Baumann, M. Peinado, and G. C. Hunt. Shielding applications from an untrusted cloud with haven. In *TOCS*, 2015.
- [10] P. Bosshart, G. Gibb, H.-S. Kim, G. Varghese, N. McKeown, M. Izzard, F. Mujica, and M. Horowitz. Forwarding Metamorphosis: Fast Programmable Match-Action Processing in Hardware for SDN. In *SIGCOMM*, 2013.
- [11] Bruce Curtis. net-tcp: TCP/IP stack bypass for loopback connections. <https://www.spinics.net/lists/netdev/msg210741.html>.
- [12] Chelsio Communications. TCP Offload Engine (TOE). <http://www.chelsio.com/nic/tcp-offload-engine/>.
- [13] B. F. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears. Benchmarking Cloud Serving Systems with YCSB. In *SoCC*, 2010.
- [14] W. J. Dally, Y. Turakhia, and S. Han. Domain-specific hardware accelerators. *CACM*, July 2020.
- [15] DDPK Authors. DDPK. <https://www.ddpk.org/>.
- [16] ETSI. Network Functions Virtualisation. http://portal.etsi.org/NFV/NFV_White_Paper.pdf.
- [17] F5 Labs. Intelligent Application traffic Management. <https://www.f5.com/products/big-ip-services/local-traffic-manager>.
- [18] N. Foster, R. Harrison, M. J. Freedman, C. Monsanto, J. Rexford, A. Story, and D. Walker. Frenetic: A Network Programming Language. In *ICFP*, 2011.
- [19] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica. Dominant Resource Fairness: Fair Allocation of Multiple Resource Types. In *NSDI*, 2011.
- [20] I. Gog, M. Schwarzkopf, A. Gleave, R. N. M. Watson, and S. Hand. Firmament: Fast, Centralized Cluster Scheduling at Scale. In *OSDI*, 2016.
- [21] Google. Flatbuffers. <https://google.github.io/flatbuffers/>.
- [22] Google. gRPC. <https://grpc.io/>.
- [23] Google. Protocol Buffers. <https://developers.google.com/protocol-buffers/>.
- [24] Google. XLA: Optimizing Compiler for Machine Learning. <https://www.tensorflow.org/xla/>.
- [25] D. Hancock and J. E. van der Merwe. HyPer4: Using P4 to Virtualize the Programmable Data Plane. *CoNEXT*, 2016.
- [26] T. Höiland-Jørgensen, J. D. Brouer, D. Borkmann, J. Fastabend, T. Herbert, D. Ahern, and D. Miller. The Express Data Path: Fast Programmable Packet Processing in the Operating System Kernel. In *CoNEXT*, 2018.
- [27] M. Jakobsson. Secure remote attestation. *IACR Cryptol. ePrint Arch.*, 2018:31, 2018.
- [28] J. Jang, S. J. Jung, S. Jeong, J. Heo, H. Shin, T. J. Ham, and J. W. Lee. A specialized architecture for object serialization with applications to big data analytics. In *ISCA*, 2020.
- [29] Jeff Barr. New - Advanced Request Routing for AWS Application Load Balancers. <https://aws.amazon.com/blogs/aws/new-advanced-request-routing-for-aws-application-load-balancers/>.
- [30] E. Y. Jeong, S. Woo, M. Jamshed, H. Jeong, S. Ihm, D. Han, and K. Park. MTCP: A Highly Scalable User-Level TCP Stack for Multicore Systems. In *NSDI*, 2014.
- [31] X. Jin, X. Li, H. Zhang, R. Soulé, J. Lee, N. Foster, C. Kim, and I. Stoica. NetCache: Balancing Key-Value Stores with Fast In-Network Caching. In *SOSP*, 2017.
- [32] A. Khawaja, J. Landgraf, R. Prakash, M. Wei, E. Schkufza, and C. J. Rossbach. Sharing, Protection, and Compatibility for Reconfigurable Fabric with AmorphOS. In *OSDI*, 2018.
- [33] D. Kim, T. Yu, H. H. Liu, Y. Zhu, J. Padhye, S. Raindel, C. Guo, V. Sekar, and S. Seshan. FreeFlow: Software-based Virtual RDMA Networking for Containerized Clouds. In *NSDI*, 2019.
- [34] P. Kumar, N. Dukkupati, N. Lewis, Y. Cui, Y. Wang, C. Li, V. Valancius, J. Adriaens, S. Gribble, N. Foster, and A. Vahdat. PicNIC: Predictable Virtualized NIC. In *SIGCOMM*, 2019.
- [35] A. Langley, A. Ridloch, A. Wilk, A. Vicente, C. Krasic, D. Zhang, F. Yang, F. Kouranov, I. Swett, J. Iyengar, J. Bailey, J. Dorfman, J. Roskind, J. Kulik, P. Westin, R. Tenneti, R. Shade, R. Hamilton, V. Vasiliev, W.-T. Chang, and Z. Shi. The QUIC Transport Protocol: Design and Internet-Scale Deployment. In *SIGCOMM*, 2017.
- [36] D. Lee, D. Kohlbrenner, S. Shinde, K. Asanovic, and D. Song. Keystone: An Open Framework for Architecting Trusted Execution Environments. In *EuroSys*, 2020.
- [37] J. Li, E. Michael, N. K. Sharma, A. Szekeres, and D. R. K. Ports. Just Say NO to Paxos Overhead: Replacing Consensus with Network Ordering. In *OSDI*, 2016.
- [38] B. M. Maggs and R. K. Sitaraman. Algorithmic nuggets in content delivery. *SIGCOMM CCR*, July 2015.
- [39] H. Mao, M. Schwarzkopf, S. B. Venkatakrishnan, Z. Meng, and M. Alizadeh. Learning Scheduling Algorithms for Data Processing Clusters. In *SIGCOMM*, 2019.
- [40] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner. OpenFlow: Enabling Innovation in Campus Networks. *SIGCOMM CCR*, Mar. 2008.
- [41] Microsoft. Fast TCP Loopback Performance and Low Latency with Windows Server 2012 TCP Loopback Fast Path. <https://docs.microsoft.com/en-us/archive/blogs/wincat/fast-tcp-loopback-performance-and-low-latency-with-windows-server-2012-tcp-loopback-fast-path>.
- [42] Microsoft. Introduction to Hyper-V on Windows 10. <https://docs.microsoft.com/en-us/virtualization/hyper-v-on-windows/about/>.
- [43] G. C. Necula. Proof-carrying code. In *POPL*, 1997.
- [44] G. C. Necula and P. Lee. Safe Kernel Extensions without Run-Time Checking. In *OSDI*, 1996.
- [45] T. Nelson, A. D. Ferguson, M. J. Scheer, and S. Krishnamurthi. Tierless Programming and Reasoning for Software-Defined Networks. In *NSDI*, 2014.
- [46] R. Nishtala, H. Fugal, S. Grimm, M. Kwiatkowski, H. Lee, H. C. Li, R. McElroy, M. Paleczny, D. Peek, P. Saab, D. Stafford, T. Tung, and V. Venkataramani. Scaling Memcache at Facebook. In *NSDI*, 2013.
- [47] E. Nordström, D. Shue, P. Gopalan, R. Kiefer, M. Arye, S. Y. Ko, J. Rexford, and M. J. Freedman. Serval: An end-host stack for service-centric networking. In *NSDI*, 2012.
- [48] E. Nygren, R. K. Sitaraman, and J. Sun. The Akamai Network: A Platform for High-Performance Internet Applications. *SIGOPS Oper. Syst. Rev.*, August 2010.
- [49] ONNX. Open Neural Network Exchange. <https://onnx.ai/>.
- [50] S. Palkar, J. Thomas, D. Narayanan, P. Thaker, R. Palamuttam, P. Negi, A. Shanbhag, M. Schwarzkopf, H. Pirk, S. Amarasinghe, S. Madden, and M. Zaharia. Evaluating End-to-End Optimization for Data Analytics Applications in Weld. In *VLDB*, 2018.
- [51] P. M. Phothilimthana, M. Liu, A. Kaufmann, S. Peter, R. Bodik, and T. Anderson. Floem: A Programming System for NIC-Accelerated Network Applications. In *OSDI*, 2018.
- [52] D. R. K. Ports, J. Li, V. Liu, N. K. Sharma, and A. Krishnamurthy. Designing Distributed Systems Using Approximate Synchrony in Data Center Networks. In *NSDI*, 2015.
- [53] D. R. K. Ports and J. Nelson. When Should the Network be the Computer? In *HotOS*, 2019.
- [54] ProxySQL. ProxySQL: A High Performance Open Source MySQL Proxy. <https://proxysql.com/>.
- [55] L. Rizzo. netmap: A Novel Framework for Fast Packet I/O. In *USENIX ATC*, 2012.
- [56] J. H. Saltzer, D. P. Reed, and D. D. Clark. End-to-End Arguments in System Design. *ACM ToCS*, Nov. 1984.
- [57] J. Sherry, D. Kim, S. Mahalingam, A. Tang, S. Wang, and S. Ratnasamy. Netcalls: End Host Function Calls to Network Traffic Processing Services. UC Berkeley Technical Report No. UCB/ECS-2012-175. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/ECS-2012-175.html>, 2012.
- [58] A. Sivaraman, A. Cheung, M. Budiu, C. Kim, M. Alizadeh, H. Balakrishnan, G. Varghese, N. McKeown, and S. Licking. Packet Transactions: High-Level Programming for Line-Rate Switches. In *SIGCOMM*, 2016.
- [59] R. Soulé, S. Basu, P. J. Marandi, F. Pedone, R. Kleinberg, E. G. Sirer, and N. Foster. Merlin: A Language for Provisioning Network Resources. In *CoNEXT*, 2014.
- [60] D. L. Tennenhouse and D. J. Wetherall. Towards an Active Network Architecture. *SIGCOMM CCR*, October 2007.
- [61] M. Walfish, J. Stribling, M. N. Krohn, H. Balakrishnan, R. T. Morris, and S. Shenker. Middleboxes no longer considered harmful. In *OSDI*, 2004.
- [62] T. Wang, H. Zhu, F. Ruffy, X. Jin, A. Sivaraman, D. Ports, and A. Panda. Multitenancy for Fast and Programmable Networks in the Cloud. In *HotCloud*, 2020.
- [63] H. Yu, A. M. Peters, A. Akshintala, and C. J. Rossbach. AvA: Accelerated Virtualization of Accelerators. *ASPLOS*, 2020.
- [64] P. Zheng, T. Benson, and C. Hu. P4Visor: Lightweight Virtualization and Composition Primitives for Building and Testing Modular Programs. In *CoNEXT*, 2018.
- [65] P. Zheng, T. Benson, and C. Hu. P4Visor: lightweight virtualization and composition primitives for building and testing modular programs. *CoNEXT*, 2018.
- [66] D. Zhuo, K. Zhang, Y. Zhu, H. H. Liu, M. Rockett, A. Krishnamurthy, and T. Anderson. Slim: OS Kernel Support for a Low-Overhead Container Overlay Network. In *NSDI*, 2019.