Oliver Michel — Research Statement

I am a networking and systems researcher working on improving the performance and scalability of *emerging networked applications*, in the face of their rapid growth, complexity, and associated resource overheads. Due to this complexity, achieving this goal warrants a holistic cross-layer approach where I cover the entire stack, from the physical layer to the higher-layer intricacies of rich, user-facing applications. My research takes a unique measurement-driven approach that emphasizes on understanding the operation and performance of applications *in the wild* and *across layers* by designing measurement methodologies and building measurement systems before crafting practical measurement studies that provide unique insight into actual deployments of large-scale systems. I use these insights to address problems across the stack by designing and building systems that leverage advances in network programmability to improve Quality of Experience (QoE) for users while simplifying and reducing overheads of running these systems for operators.

Research Vision. Many aspects of our lives, including in healthcare, education, and e-commerce, rely on increasingly sophisticated networked applications, like video conferencing and virtual reality, while the applications themselves critically depend on the performance of the underlying network. Operating and scaling these applications is hard as they evolved into *complex and resource-intensive systems* associated with *stringent performance and predictability requirements*. Not meeting these requirements affects all stakeholders: Users may experience poor QoE, such as choppy video calls, while operators are faced with significant revenue loss for just milliseconds of added latency or seconds of downtime. The ultimate goal of my research is to solve this conundrum and to enable scalable and high-performance modern networked applications. Toward this goal, my work is driven by two high-level challenges, which I outline next.

Closing the Semantic Gap between Applications and Networks. My measurement studies have revealed a common theme: the lack of transparency and coordination between applications and networks. Autonomous, closed-loop network control promises to enable such coordination with the goal of optimizing networks for modern applications. Existing network-level monitoring systems, however, do not provide the required application-level insight to realize this vision. This results in networks that do not understand the relevant metrics and performance objectives of applications and, consequently, are unable to optimize themselves for the applications' needs. Conversely, applications often see the network through a clouded window and are unable to correctly interpret artifacts from the network (e.g., sudden latency spikes). To solve this problem, I design measurement methodologies and systems and craft measurement studies that *provide the required application-level insights* to enable closed-loop network control at scale [4, 5, 7, 8].

Boosting Application Performance and Scalability. With surging demand, growing link capacities, and the end of Moore's law, it is increasingly difficult to efficiently run and scale networked systems, posing a significant problem for operators and ultimately compromising QoE. This affects both the class of emerging applications described above as well as lower-level, packet-processing systems such as firewalls or routers. In my work, I show through building systems how leveraging application-level monitoring data and insights inside the network can facilitate significant performance and quality-of-service improvements for a wide range of applications. The resulting networked systems ultimately *tie the network and applications* and use a mix of traditional optimization techniques within novel system architectures [1, 2] as well as modern network programmability using accelerators and programmable switches [3, 4, 6].

My measurement-driven approach tackling these two fundamental challenges of modern networked systems is fruitful as it enables a natural workflow where bridging the semantics of applications and networks reveals cross-layer optimization opportunities, leading to more powerful, reliable, and scalable systems. My work has provided important results, brought new insights, and improved system and application performance. More broadly, however, it has had practical impact on the networking community: methods and artifacts of my work have been adopted by industry, e.g., at Arista Networks, and are widely used by Universities, e.g., at Columbia University and the University of California Santa Barbara.

Outline – My research spans three areas. First, my work on *video conferencing* shows how to measure and optimize complex, emerging applications. Second, my work on *Linux networking* takes a more general approach and explores how to accelerate a wide range of network functions. Third, my work on *network monitoring and control* combines learnings from measurement and system design, enabling closed-loop network control.

Performance and Scalability of Video-Conferencing Applications

My work on video conferencing started with enabling visibility into the increasingly important area of real-time communications, which also includes, for example, live streaming and cloud gaming. These systems have stringent performance and scalability requirements which are often hard to meet in today's networks and using today's implementations. My research here incorporates both better understanding these applications in different settings through new measurement approaches, as well as, improving their performance and scalability through novel system designs for the underlying infrastructure driving these systems today.

Enabling Passive Measurement of Zoom Performance. To better understand and manage video-conferencing applications, we need effective ways to measure their performance in the wild to aid operators in capacity planning, troubleshooting, and setting QoS policies. Previous large-scale measurements of these applications rely on end-host cooperation, which is impractical for network operators, while in-depth analysis of captured packets requires knowledge of the packet formats. Zoom is one of the most sophisticated and popular applications, but it uses a proprietary, previously not understood network protocol. In this work [5], we were the first to demystify how Zoom works at the packet level and propose techniques for analyzing Zoom performance solely from passively captured packets, revealing important metrics like media bit rates, frame rates, latency, and jitter. This required extracting the structure and hierarchy of media frames, media streams, and ultimately meetings from millions of packets, despite encryption, the unknown binary packet format, and dynamic port mappings. Finally, we provide unprecedented, at-scale insights into Zoom's real-world operation using packet-level data for every call on the entire Princeton campus. Our study shows a long tail of poor meeting performance across several key quality metrics, such as jitter and bit rate. All software artifacts are available for the community and have been adopted by industry and academics to study Zoom and other video-conferencing systems in the wild.

Redesigning Video-Conferencing Systems for Scale. Our measurements of Zoom and other applications have revealed that the architecture of video-conferencing systems today is not well suited for the current and expected growth of these applications, which require an unprecedented scale of their complex underlying infrastructure. This, in particular, concerns the servers that relay media among participants which must sustain high data rates and dynamically scale compute and networking resources up and down while also supporting complex logic for rapid adaptation of media streams. We found that this workload is inherently hard to scale as the amount of packets to process grows quadratically with meeting size and under-provisioning of these servers has direct impact on QoE. Our analysis of Zoom, however, also revealed that forwarding and adapting media signals in video conferencing is strikingly similar to traditional packet processing. Based on this insight and following a long line of work on software-defined networking (SDN), we show how to rearchitect video-conferencing systems and offload all media forwarding to an efficient *data plane* which runs on line-rate programmable switches, increasing throughput by 20× to 40× compared to a commodity server [6]. Our prototype also lowers latency and jitter by roughly 10× in comparison to software which are key quality metrics for real-time applications.

Seeing and Mitigating Wireless Impact on Video Conferencing and Beyond. Rapid delay variations in today's wireless networks impair the QoE of low-latency, interactive applications, such as video conferencing. Our measurements have revealed that applications essentially operate in the dark as it is hard to differentiate between short-term delay artifacts stemming from wireless scheduling and retransmissions and those caused by congestion. To tackle this problem, we propose Athena [8], a framework that correlates high-resolution measurements across Layer 1 to Layer 7 to remove the fog from the window through which today's congestion-control algorithms see the network. Athena extracts 5G control-channel telemetry of physical-layer data units, retransmissions, and scheduling decisions and precisely time-synchronizes these measurements with packets at the network layer and video frames or audio samples at the application layer. As a result, Athena can identify the cause of distorted audio or stuttering video and feed this information back to the applications. This cross-layer view of the network empowers the networking community to revisit and re-evaluate scheduling and congestion-control algorithms in light of the complex, heterogeneous networks that are in use today, paving the way for network-aware applications and application-aware networks.

Linux Networking and Acceleration of Network Functions

In other works [3, 4, 6, 7], I have accelerated different packet-processing and network-management tasks using hardware offload. Yet, we need general approaches for accelerating a wide range of network functions automatically and transparently. Linux is the most commonly used platform for software packet processing but supporting today's required performance with its out-of-the-box networking stack is challenging.

Transparently Accelerating Linux Networking. In my work on software-based packet processing, building on years of research on high-performance software data planes, we introduce transparent acceleration into the Linux networking stack [1, 2]. Rather than using kernel-bypass technologies such as DPDK or XDP as an alternative pipeline, which requires re-implementation of protocols and custom control-plane interfaces, we leverage eBPF technology in a novel way, as an explicit fast path for common packet-processing tasks, such as routing, packet filtering, and intrusion detection. We integrate this fast path with the kernel's existing control plane, unifying state between fast and slow paths. In this innovative design, standard command-line tools, control-plane software, and frameworks such as Kubernetes still work without modification and benefit from a faster network data plane with throughput improvements of up to 77% for various workloads.

Real-Time Network Telemetry and Control

My work on network telemetry focuses on the question of whether we can obtain and effectively leverage insights gathered from hundreds of millions of packets per second in today's networks. This is important for security applications but also enables novel congestion-control algorithms. By taking a holistic approach that partitions measurement and analytics between hardware and software, we show that this goal is achievable.

Packet-Level Network Telemetry and Analytics with *Flow and Jetstream. My first line of work in this area explored whether telemetry systems can harness the performance of programmable line-rate switches while also meeting requirements for practical deployment and efficient integration with analytics and security platforms. We introduce *Flow [4], a practical telemetry system that is highly efficient and supports concurrent measurement and dynamic queries through carefully partitioning between hardware and software. The pipeline operates at 3.2 Tbit/s and exports a stream of records that provides unprecedented packet-level insight into the network and can be ingested by a wide range of applications, e.g., for intrusion detection or network debugging. *Flow is accompanied by a specialized, high-performance software packet-analytics engine called Jetstream which is able to process up to 250 million packets per second on a commodity server, a significant improvement of two orders of magnitude over previous systems [7].

PowerTCP: Telemetry-Assisted Network Congestion Control. Especially in the context of data centers where congestion control algorithms must act on sub-millisecond time scales and rapidly adapt to changing network conditions (e.g., frequent micro bursts), leveraging telemetry is a promising way to provide insight into the queue state and current network congestion. In PowerTCP [3], we designed an innovative end-host congestion-control algorithm that, in contrast to the state of the art, operates on both network state (e.g., absolute queue length and round-trip time (RTT)) and network variation (e.g., queue length and RTT gradient) with a data-plane component that provides telemetry about bottleneck queue occupancy in real time to TCP senders. PowerTCP is able to utilize available bandwidth within one or two RTTs while being stable, maintaining low queue lengths, and resolving congestion rapidly, making it particularly suitable for reconfigurable data-center networks where the link capacity changes frequently.

Future Research Directions

I am excited to pursue more research in the future that is motivated by the challenges modern networked applications face. My measurement-driven approach has uncovered various complex and often unintended ways networks and applications interact, such as the impact of latency artifacts from access networks on video-conferencing applications. Secondly, I have shown that cross-layer designs and optimizations are effective in mitigating many of the challenges that applications face today.

Going forward, I plan to leverage these insights to work on pressing issues such as building networks that provide the best possible service to the applications they carry. We live in an exciting time where this goal is more achievable than ever through network programmability: the network is no longer a black box but can be programmed and optimized, for example using programmable switches or programmability inside the cellular core. Furthermore, I plan to work on more forward-looking directions, such as future network designs that provide the right abstractions to facilitate telemetry and information sharing across layers. Below, I present three specific directions.

Video-Conferencing Rate Control for Wireless Networks. Our measurements enabled by Athena [8] have underlined that wireless networks are a challenging environment for real-time applications. This poses a problem as our measurements on Zoom revealed that the majority of video-conferencing traffic on our campus is carried over at least one wireless access network. Through a combination of link-layer and transport-layer measurements, we found that scheduling and retransmission mechanisms in cellular networks introduce delay artifacts that are not due to congestion but mislead today's predominantly delay-based congestion-control algorithms. Athena can feed real-time data about the root cause of individual artifacts to congestion controllers. Our early results show that then redesigning congestion-control algorithms to differentiate between congestion-induced delay and other delay artifacts is promising in dramatically boosting QoE for users.

Networks Optimized for Real-Time Applications. I am currently working on applying lessons learned from our work on systems for network telemetry and measurement of video-conferencing applications to build practical systems for continuously analyzing the QoE of a wide range of interactive applications and using this information to control network devices. The main measurement challenge is to efficiently analyze the complex semantics of these applications in real time in the data plane and to identify the root cause of degraded quality. Our work on Zoom monitoring is the foundation of this and our work on offloading video-conferencing infrastructure proposes building blocks to realize such an in-network monitoring system [5]. The second set of challenges arises from identifying the right control actions. Options here include intelligent dropping of packets that are less relevant for QoE or scheduling transmission grants in cellular networks in a way that is synchronized with the applications' media streams. Both my work on a data-plane based SFU [6] and Athena [8] enable this direction.

Enabling Cross-Layer Visibility in the Network Stack. Looking further into the future, I aim to develop more principled ways to share critical information across layers in the networking stack. The interfaces between layers today allow passing data up and down the stack but are largely not designed to carry meta data relevant for decision making at adjacent layers. Consequently, we require abstractions that enable such information sharing for cross-layer optimization while leaving the layering abstraction intact. For this, I plan to collaborate with industry partners and standardization bodies to design future-proof abstractions that are suitable for a wide range of use cases. The design space for such abstractions is vast and includes options ranging from in-band mechanisms (e.g., L4S for congestion control) to more complex data structures that are shared across layers out of band.

References

[1] Marcelo Abranches, Erika Hunhoff, Rohan Eswara, Oliver Michel, and Eric Keller. Linuxfp: Transparently accelerating linux networking. In *IEEE ICDCS*, 2024.

[2] Marcelo Abranches, Oliver Michel, and Eric Keller. Getting back what was lost in the era of high-speed software packet processing. In *ACM HotNets*, 2022.

[3] Vamsi Addanki, Oliver Michel, and Stefan Schmid. PowerTCP: Pushing the performance limits of datacenter networks. In *USENIX NSDI*, 2022.

[4] John Sonchack, Oliver Michel, Adam J. Aviv, Eric Keller, and Jonathan M. Smith. Scaling hardware accelerated network monitoring to concurrent and dynamic queries with *Flow. In *USENIX ATC*, 2018.

[5] Oliver Michel, Satadal Sengupta, Hyojoon Kim, Ravi Netravali, and Jennifer Rexford. Enabling passive measurement of Zoom performance in production networks. In *ACM IMC*, 2022.

[6] Oliver Michel, Satadal Sengupta, Hyojoon Kim, Ravi Netravali, and Jennifer Rexford. Rearchitecting video-conferencing for scalability using SDN principles, 2024. *In submission.*

[7] Oliver Michel, John Sonchack, Eric Keller, and Jonathan M. Smith. Packet-level analytics in software without compromises. In *USENIX HotCloud*, 2018.

[8] Fan Yi, Haoran Wan, Kyle Jamieson, Jennifer Rexford, Yaxiong Xie, and Oliver Michel. Athena: Seeing and mitigating wireless impact on video conferencing and beyond. In *ACM HotNets*, 2024.