Research statement

Ankit Singla

February 15, 2021

My objective is to pursue **ambitious, high-risk, long-term research** on network design. My work uses experimental and analytical methods to explore novel network infrastructures, algorithms, and protocols. I strive to frame provocative research questions that either re-examine what is seen as settled wisdom, or break new ground in creative, under-explored research spaces. I discuss my **three most significant sets of results**, followed by two smaller yet important focus areas, and close with a summary of my research group's track record by the numbers. Throughout, I discuss only my work over the **past 5 years**, since January 2016, when I started at ETH.

1 Networking with massive-scale low Earth orbit constellations

[ACM IMC 2020 — Best Paper] [ACM CoNEXT 2019 — IRTF Applied Networking Research Prize]
[ACM HotNets 2020] [ACM HotNets 2020] [ACM SIGCOMM CCR 2020] [ACM HotNets 2018]
Two other pieces of work are under review, with the submitted drafts available on request.

My group is leading this nascent research area, driven by the rapid industry developments, with SpaceX having already deployed 1000+ satellites. Such constellations promise global Internet connectivity, and for long-distance routes, much lower latency than today's fiber paths. However, fully exploiting their potential requires tackling their inherent challenges: thousands of low-Earth orbit satellites travel at high velocity relative to each other, and relative to fixed terrestrial ground stations. Our work explores the **implications of the resulting highly-dynamic connectivity, showing how the Internet's core design primitives, which assume a largely static core infrastructure, are inadequate for such connectivity.** At every layer of the network stack, our work quantitatively explores the new challenges that arise in this setting, together with novel solutions to tackle them:

- Our work makes the case for analysis tools specialized for satellite networks' dynamics, and to fulfill this need, we built Hypatia, a packet-level simulation and visualization framework for such networks. We also learnt several insights from Hypatia, namely that satellite motion manifests in changes in path structure and end-end latency, which, in turn, break standard primitives for congestion control and traffic engineering. For example, delay-based congestion control is limited on such networks because delay changes do not reliably reflect queuing. Loss-based congestion control is also undesirable — waiting for loss implies increased delay from queuing, counter to the low-latency promise of these networks. This result holds deep significance — **our main end-end signals for congestion control,** *i.e.,* **loss and delay, are both defective here**. "Easy" fixes, like using the predictability of the satellite dynamics to account for them explicitly, would require casting aside the end-end principle, and be unworkable with modern protocols like QUIC. **[ACM IMC 2020 — Best Paper]**
- Satellite motion results in time-variant connectivity from the satellite network to ground stations, especially at early stages of deployment. We argue that naively using the Internet's BGP protocol to connect such networks to terrestrial Internet service providers results in either sub-optimal use of the intermittent connectivity, or routing failures (*e.g.,* due to BGP disabling time-variant routes to prevent "route flapping"). We design **a scheme for integrating such networks into the Internet**, overcoming these issues. **[ACM SIGCOMM CCR 2020]**
- We show that for the problem of deciding how to interconnect satellites to each other, traditional design methods are ineffective due to the transient visibility of satellites from each other. Leveraging the symmetry and geometry of such constellations, we define a **family of topologies from which it is easy to identify latency or throughput optimized networks**. Our approach improves substantially (by roughly 50%) over the *de facto* baseline topology. **[ACM CoNEXT 2019 — IRTF Applied Networking Research Prize]**
- Instead of determining a static placement of network routers, we need to determine what satellite trajectories to use, such that the network meets its performance goals. Besides exploring the optimization of satellite trajectories while accounting for rich domain-specific constraints (such as the impact of launch site locations), our work also sheds light on what objectives the planned constellations are optimizing for. **[Work under review]**

Summary and future plans: Besides **two award papers**, this work has also received substantial press coverage, with at least 9 news articles and a technical blog. A listing of these is maintained here: https://bdebopam.github.io/. This work is **full of "firsts"**, concretely drawing out the numerous challenges involved in this emerging area, providing first-cut solutions to some of them, and supplying the tools to make progress on others — I expect our Hypatia framework to become the **default experimentation platform** for such networks. To ensure sound footing with the right assumptions for our work, my group has drawn on our conversations with **Astrome**, an Indian space-sector startup hoping to launch its own satellite network. For greater industry and academic participation, following up on the IRTF's award of its Applied Networking Research Prize to our work, I have initiated a conversation with the IRTF about starting an **IRTF research group on satellite networking**.

I am also drawing up plans for a cross-university experimental platform, **SatNetLab** (inspired by PlanetLab) with user terminals from SpaceX and other companies hosted at universities, allowing researchers to conduct real experiments to test out ideas, such as for congestion control and at the application-layer, on the real operational satellite networks. My ongoing evangelism of SatNetLab is quite successful — networking faculty at 5 **top institutions** have expressed interest in participating. In terms of research directions, besides searching for solutions that improve on our first set of results above, we are also exploring new problems like the resilience of such networks to denial-of-service attacks, and routing and traffic engineering within the constellation.

2 Data centers thread A: the capabilities and limits of optical networking

[ACM HotNets 2016 – single author work] [ACM SIGCOMM 2017] [ACM SIGCOMM 2020]
One more piece of work is under review; the submitted draft is available on request.

The problem of how to interconnect data center servers in a cost-efficient and performant manner is a large subarea of networking. My work since joining ETH specifically tackles the **role of optical networking in this setting, deeply examining its value and its limits**. I have the following three highlight results on this topic:

- A decade-long research thread suggested a variety of optical technologies to dynamically connect servers that need high-bandwidth connectivity. The intuition here is appealing: instead of building for the worst-case traffic with an expensive electrical fabric throughout the data center, we should dynamically establish connections to respond in real-time to demand where it arises. However, my work rigorously shows that given the technology costs, this intuition does not pan out — such dynamic networks are inferior (when compared fairly at equal cost) to static expander-graph-based electrical networks. This is a **surprising result: dynamic networks do not win out over static ones**, especially for the setting from which this intuition stems, *i.e.,* when traffic is skewed, with a few traffic hot-spots. **[ACM HotNets 2016 – single author work] [ACM SIGCOMM 2017]**
- Cloud infrastructure operators face huge demand for computing in the vicinity of large metro areas. However, in such areas, there are very limited options for sites for mega-scale data centers. In a collaboration with Microsoft, we thus explore how to interconnect a set of smaller data centers located within tens of kilometers of each other. This work is the **first to flesh out and propose a solution to the problem of designing regional data center network interconnects**. Beyond the newness of the problem and its tradeoff space, the conclusions are also surprising: neither per-packet electrical switching, nor fine-grained optical wavelength switching are well-suited for such networks — instead, coarse-grained optical networking (switching entire fibers using optical space switches) strikes the right cost, complexity, and performance trade-offs. **[ACM SIGCOMM 2020]**
- We may be hitting the limits of scalability of cheap commodity electrical switches, especially given the ever higher per-port line rates, *e.g.,* a switch capacity of 25.6 Tbps translates to only 32 ports at 800 Gbps. Low-port-count switches make inefficient networks, as each end-end path traverses many switches, increasing latency and decreasing network bandwidth. Unfortunately, this issue cannot be addressed by the large body of work on data center optics that still relies on using high-port-count electrical switches at the top-of-rack level. My work, in collaboration with researchers at HUJI, tackles this with a novel network design that equips each *server* with optical network interfaces, such that most of the network operates fully in the optical domain. A parallel low-capacity electrical network serves to bootstrap connectivity and transition large flows to the optical network. To the best of my knowledge, this is the **first exploration of server-based optics**. **[Work under review]**

Summary and future plans: I have produced both work that deeply critiques a long thread of data center research, and work that starts off new directions addressing the pressures facing the industry. Despite my long experience in this area, there are still interesting blind spots I'm discovering and exploring. One instance is the problem of designing data center networks explicitly for small scale (say, up to ten thousand servers) — *most data centers are in fact of this size*, but research tends to focus on the mega-scale data centers run by a handful of companies. In my collaboration with HUJI, I'm investigating for this setting, a variety of design techniques that would be typically dismissed without serious consideration as "not scalable" due to the focus on mega-datacenters.

3 Data centers thread B: workload-aware scheduling

[ACM SoCC 2020] [ACM SoCC 2020] [ICML 2019] [USENIX HotCloud 2019] [USENIX NSDI 2019]

My work in this direction is about the cloud operator obtaining implicit and explicit information about their workload characteristics, and leveraging this information in scheduling to improve efficiency.

- Our work breaks new ground in **partial-knowledge scheduling**, including: (a) a thorough analysis of various methods of obtaining workload information (specifically, network flow size when a flow starts) useful in packet and flow scheduling; (b) concretely arguing that in most settings this information will be incomplete and imprecise, and thus the scheduling contexts of greatest interest are ones with partial workload knowledge; (c) revealing, counter-intuitively, that in such contexts, common scheduling algorithms sometimes result in degraded performance when the amount of workload knowledge available increases; and (d) proving for a certain packet scheduling algorithm, this guarantee of monotonically non-deteriorating performance as more effort is invested to obtain workload knowledge. The last two results here have **deep implications for the incentives for tenant applications** — if there are no guarantees that their performance will improve, or at least not deteriorate, then there is limited incentive to provide workload information. **[USENIX NSDI 2019]**
- New computing platforms and emergent applications create new opportunities for optimizing scheduling. For instance, in serverless computing contexts, our work observes the large overheads paid for setting up the context for a task, and proposes that this cost be shared across multiple concurrent executions of the same task **[ACM SoCC 2020]**. Our work also addresses how one can tailor network packet scheduling for common communication patterns like partition-aggregate or scatter-gather, using careful per-packet control afforded by a hardware offload strategy **[ACM SoCC 2020]**. We take a new and different approach in our treatment of emergent machine learning training workloads — we observe that many such workloads do not need a "guaranteed delivery" primitive from the network, and we can relax this constraint to benefit other applications with more stringent needs. **[ICML 2019]**
- This work is as much about interfaces and abstractions, as it is about optimization — enabling context- and workload-dependent scheduling optimizations requires making the infrastructure aware of these dependencies. We thus explore **minimal and simple-to-use interfaces that still provide enough information to the infrastructure to enable better scheduling, without imposing a large burden** on either the infrastructure or the tenant applications. **[ACM SoCC 2020] [USENIX HotCloud 2019]**

Summary and future plans: We plan to explore new degrees of flexibility and new constraints in emerging scheduling contexts like serverless computing, *e.g.*, how to make smart-yet-fast scheduling decisions for workloads that themselves only last milliseconds? After all, scheduling should not incur time that is comparable to a workload's execution time. We are also exploring primitives for "approximate networking", where we can exploit the tolerance of certain workloads to deteriorated network performance, to benefit workloads with more stringent expectations: how do we guarantee only *bounded degradation* when we relax network service primitives?

4 Smaller focus areas around latency and user experience over the Internet

I next discuss two areas, where we have smaller yet exciting and important bodies of work.

4.1 New algorithmic opportunities in video streaming

[USENIX ATC 2020] [PAM 2020]
One more piece of work is under review; the submitted draft is available on request.

One of my goals is to understand better how today's popular streaming platforms operate. This work involves tailoring measurement scenarios to yield useful data, and then analyzing this data manually or using automated means. A key result so far is that the behavior of the streaming rate adaptation algorithm of popular services can be reconstructed in an automated fashion with high accuracy with only blackbox interactions with said services over the Internet. Further, this can be done in such a way that the code generated is both concise and understandable to those with exposure to common primitives used in such algorithms. Our work also shows that using machine learning tools out-of-the-box is insufficient to achieve these objectives: one loses either accuracy or conciseness. This result has significant implications, namely, that **proprietary logic developed at great expense can be relatively easy to reconstruct if used in public-facing Internet services**. We are working on applying a similar approach to congestion control, but this requires addressing new challenges related to long-term state that congestion control algorithms sometimes use. **[USENIX ATC 2020] [PAM 2020]**

Our analysis of existing platforms has also revealed new opportunities for improving video streaming, by jointly considering the offline (video encoding and chunking) and online (playback) phases of streaming. Specifically, we found that while the most sophisticated platforms already use complex video coding pipelines with custom bitrate design for different content, different scenes, or even different shots within a scene, these customizations are performed independently of their later interactions with online rate adaptation. Our work shows how this can deteriorate performance under challenging network conditions, and suggests **novel ways of joint optimization of the offline and online phases to improve the viewing experience for stream consumers**. We are collaborating with a large streaming service to explore if this strategy could be effective on their platform.

4.2 The Internet at the speed of light

[ACM IMC 2020] [PAM 2020 – Best Dataset Award] [HotCloud 2017] [PAM 2017 – Best Dataset Award]
[Arxiv 2018] [Arxiv 2018]

The pursuit of nearly the lowest possible latency for long-distance communication over the Internet is a 'grand challenge' project for me. The physical lower bound is the *c*-latency, the time it takes to traverse the distance between the communicating end points along the geodesic on Earth's surface at the speed of light in vacuum. How far are we from this limit, and how do we get closer? Our measurement work has shed light on why the Internet is too slow **[PAM 2017 – Best Dataset Award] [Arxiv 2018]**, how this impacts applications like advertising **[PAM 2020 – Best Dataset Award]**, and how the high-frequency trading industry builds infrastructure to achieve nearly *c*-latency in a limited niche context **[ACM IMC 2020]**. Applying the lessons from these measurements, we are developing a proposal to augment the Internet with a nearly-*c*-latency infrastructure **[Arxiv 2018]**.

5 A summary by the numbers, since my January 2016 start at ETH

- 25 peer-reviewed publications, including …
  - 12 papers at top-tier conferences: IMC 2020 ×2, SIGCOMM 2020, SoCC 2020 ×2, ATC 2020, SC 2020, NSDI 2019, CoNEXT 2019, ICML 2019, SIGCOMM 2017, and SC 2016.
  - 5 direction-setting papers at the premier networking workshop, HotNets, including one as the sole author. **HotNets'** 22% **accept rate over the years is as competitive as flagship conferences in other areas**.
- 5 awards: Best Paper at IMC 2020, IRTF's Applied Networking Research Prize for our CoNEXT 2019 work, Best Dataset at PAM in both 2017 and 2020, and a Google Faculty Research Award for 2018.
- At least 12 articles about our work appeared in the news or on popular academic blogs.
- Of the two top networking conferences, NSDI and SIGCOMM, I served on the program committee 8 times out of the 10 total possibilities in these years.
- I advise 7 PhD students, two of whom, Debopam Bhattacherjee and Vojislav Dukic, will graduate in 2021.
- Over this period, I have collaborated with Microsoft, Mellanox, Disney, Swisscom, ABB, and Huawei.