The Skillful Interrogation of the Internet

Mark Crovella
Boston University
Department of Computer Science
crovella@bu.edu

This article is an editorial note submitted to CCR. It has NOT been peer reviewed.
The author takes full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

As SIGCOMM turns 50, it's interesting to ask how networking research has evolved over time. This is a set of personal observations about the "mindset" associated with Internet research.

In reflecting on SIGCOMM's 50th anniversary, I want to write about finding a well-motivated approach to the study of the Internet. The first thing to ask is: what is the general mindset one should take when thinking about Internet study? If you open a textbook on networking, you will find that networking's mindset, its implied purpose, is to solve communication problems by building networks based on strong engineering principles. For example, you will find discussion of how to ensure reliable packet delivery or how to compute efficient routes in a distributed fashion.

There is a sort of single-agency mindset in this view of networking. It suggests a world in which network engineers working together design and deploy a well-understood, optimized system. But this mindset does not go very far in helping us understand the Internet today. The construction of the Internet is basically outside this model in reality. The construction of the Internet is the result of complex interplay between engineering strategies, societal demands, public policy, commercial activity, business competition, and latent drivers like geography and levels of economic development. And furthermore, there is a strong feedback loop that exists between the Internet and society: just as it's very hard to build computers without using other computers, it would be hard to build the Internet from scratch if we didn't have a powerful network already to coordinate all the activity needed.

The multi-agent nature of Internet construction means that much of the Internet is hidden in one way or another. A reviewer of this article pointed out how much of the Internet's infrastructure is not often described in a typical networking textbook: for example content delivery networks, as well as middleboxes for load balancing, compression, caching, traffic shaping, encryption, firewalling, and network address translation. These devices touch a very large amount of Internet traffic. Likewise, a huge part of the Internet is hidden behind corporate firewalls.

As a result, at a certain point, researchers began to implicitly acknowledge that to understand this enormous complexity required a fundamental shift in mindset. For me, this started when a friend first showed me how to use traceroute to debug an unresponsive application. I subsequently got this sense really strongly when reading Jean Bolot's 1993 SIGCOMM paper on experimental observation of packet queueing [2] (which was a big inspiration for our development of bandwidth measurement tools). Another early paper with a similar mindset is [1], which is the subject of a retrospective in this volume.

It's been quite interesting to me to observe the tendency that eventually emerged, which is essentially to treat the Internet as a natural phenomenon to be investigated. That is, network measurement researchers often behave like biologists or social scientists. Researchers study the Internet as a biologist would study a newly discovered organism. This is a conceptual shift away from reasoning only from first principles, towards a fusion of engineering knowledge with empirical discovery. It's been really intriguing to me that although the Internet is a 100%-human-constructed artifact, it is often treated in some ways as if it were a newly discovered species or cell type.

I have to confess that although I was involved in Internet measurement from its early days, I sometimes used to wonder whether the Internet measurement field was "worthy" of distinct definition – was it a well-defined intellectual pursuit? I eventually realized that indeed, Internet measurement certainly is intrinsically distinct. This is in part because it requires this shift in mindset toward regarding the Internet as a quasi-natural phenomenon, a shift away from an "engineering" and toward a "scientific" mental posture.

Another important reason is that an empirical approach to understanding the Internet requires great effort in the development of tools and methods. In her biography of R.A. Fisher [3], his daughter Joan Fisher Box wrote:

> The whole art and practice of scientific experimentation is comprised in the skillful interrogation of Nature.

Skill is particularly required because Nature

> ...responds to the form of the question as it is set out in the field and not necessarily to the question in the experimenter's mind; she does not interpret for him; she gives no gratuitous information; and she is a stickler for accuracy.[1]

When I read these lines I am reminded strongly of the many ingenious tools and insightful methods that researchers have developed in pursuit of knowledge about the Internet. For better or worse, the Internet does not give up its secrets readily.

As well, this style of investigating the Internet means that the analytic tools needed can be like those used by natural and social scientists. Internet research often involves statistical hypothesis

[1] A thought-provoking discussion of this passage (and more) is presented in [4, Ch. 4].

testing, parametric and nonparametric tests, causal analysis, controlled experiments – even randomized controlled trials. This distinct set of tools tends to set Internet measurement research apart from networking research more generally. For example, it's quite noticeable how important statistical inference is in Internet measurement research. Of course, inference is such an important tool because the Internet is so good at hiding its properties, as discussed above.

And finally, of course, the intrinsic significance of Internet research is heightened by the need to understand the Internet due to its immense impact on society.

So, I now see Internet research as an example of the "skillful interrogation of Nature." What has resulted from this shift in perspective?

This shift in thinking about the Internet from an "engineering" to a "scientific" posture has led to a raft of important observations, surprising invariants, and useful rules of thumb. It has led to discoveries such as network traffic self-similarity, the complex topology of the Internet, and the prevalence of heavy-tailed distributions in Internet measurements. As an aside, it's interesting to note that generally, heavy-tailed distributions are associated with a wide variety of phenomena in economics, social science, and the natural world, and only occasionally with engineering.

This shift in thinking has also led to Internet research becoming a very fertile ground for data science. And none too soon, because a significant amount of progress in network security has arrived through data analysis. Nowadays, security in depth is impossible without sophisticated analysis of network measurements.

Looking forward, I urge networking researchers to view their field as having an important scientific and discovery-based component. Further, I hope that networking research will seek greater explanatory power. Over time, the field of biology has progressed from a mainly-descriptive science to one that primarily seeks mechanistic explanations for natural phenomena. Darwin led the way, by making a conceptual leap suggesting that the causes of natural diversity might be comprehensible. This progress been incredibly challenging, and has required discoveries ranging from evolution to biochemistry to genetics. I think Internet research could learn a lot by pushing toward similar goals. I am not suggesting a return to reasoning from first principles, but rather an effort to elicit what those first principles are in a broader setting. What principles dictate the nature of network engineering? Can we understand what factors drive the development and deployment of network infrastructure? What drives the need for digital, networked communication? How do society and communication networks feed on each other?

These questions suggest to me that the scientific approach to Internet research is only beginning. I am excited to continue to see and participate in the progress of that research as a part of the skillful interrogation of Nature.

ACKNOWLEDGMENTS

I appreciate the comments of the reviewers for this special issue, which improved this article. More generally I am deeply grateful to my colleagues and students over the years who have worked with me in adopting this viewpoint towards Internet research.

REFERENCES

[1] D. R. Boggs, J. C. Mogul, and C. A. Kent. 1988. Measured Capacity of an Ethernet: Myths and Reality. In *Symposium Proceedings on Communications Architectures and Protocols (SIGCOMM '88)*. ACM, New York, NY, USA, 222–234. https://doi.org/10.1145/52324.52347

[2] Jean-Chrysotome Bolot. 1993. End-to-end Packet Delay and Loss Behavior in the Internet. In *Conference Proceedings on Communications Architectures, Protocols and Applications (SIGCOMM '93)*. ACM, New York, NY, USA, 289–298. https://doi.org/10.1145/166237.166265

[3] J.F. Box. 1978. *R. A. Fisher, the life of a scientist.* Wiley.

[4] Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect* (1st ed.). Basic Books, Inc., New York, NY, USA.