10 Networking Papers: Elegance and Insight

Venkata N. Padmanabhan
Microsoft Research India, Bangalore

When Jim Kurose invited me to write a piece for the "recommended reading" series in CCR, I thought it would be a fun exercise and accepted the invitation right away. Little did I realize how challenging it would be to put together this list, as much because I was only allowed to pick 10 papers as because I had to steer clear of the many fine papers that had already been picked in previous lists. It has nevertheless been an enjoyable experience and I thank Jim for providing me with this opportunity. I also thank him, Zahir Koradia, and Ram Ramjee for their comments on this article.

Rather than focusing on a single topic, I decided to pick papers from across sub-areas of networking that have left a lasting impression on me (and quite possibly on other researchers as well) because of their elegance, the insights they provide, and in many cases both of these. I hope many readers will enjoy reading these papers, if they have not already done so.

Now let me get started with the list. Adaptation is a key aspect, indeed requirement, of any networked protocol or system. The first few papers in my list focus on various aspects of adaptation. We will start with:

- D. Chiu and R. Jain, "Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks", *Computer Networks and ISDN Systems, Vol. 17, No. 1, June 1989, pp. 1-14.*

Congestion control for unicast flows is one problem where adaptation has been applied with great success, spurred by the congestion collapse episodes in the Internet in the mid-1980s. Many readers will be familiar with the DECBit scheme by Ramakrishnan and Jain [6] (the precursor to the modern-day explicit congestion notification (ECN) [5] mechanism for TCP/IP) and the TCP congestion control algorithm by Jacobson [3]. The paper by Chiu and Jain provides a really elegant argument for why additive increase and multiplicative decrease (AIMD) is the right control mechanism in such schemes, for achieving efficiency and fairness. The graph schematics in the paper provide nice intuition for how the various increase and decrease policies operate.

How about congestion control for multicast traffic? This brings us to our next paper:

- S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast", *ACM SIGCOMM 1996, Stanford, CA, USA, August 1996.*

There are a couple of ways in which congestion control for multicast differs from that for unicast. First, there is no single "correct" rate to send at, given the bandwidth heterogeneity across the receivers. Second, it would be unscalable for a sender to receive and process congestion feedback from a large number of receivers. The work by McCanne et al. addresses both of these issues with a simple idea — make the receivers responsible for congestion control. In RLM, the sender stripes layered information (e.g., layered video) across multiple multicast groups and each receiver adapts to *its* dynamically varying bandwidth by adding or dropping layers. While this is an elegant framework, it seems to have been ahead of its time and also been held back by the requirement that the source data be amenable to layering. In practice, manual adaptation appears to be the norm for multicast, e.g., with users choosing from a few bandwidth options when tuning in to a webcast.

Network routing is another area where adaptation has been applied with success. In particular, routing protocols are designed to adapt to link failures. But how about adapting to changes in link load? Our next paper considers this issue:

- A. Khanna and J. Zinky, "The Revised ARPANET Routing Metric", *ACM SIGCOMM 1989, Austin, TX, USA, September 1989.*

This paper talks about the instability caused by the delay-based routing metric used in the ARPANET. It then goes on to present a revised metric that behaves similar to a delay-based metric under light load and to a more stable, capacity-based metric under heavy load. Nearly 20 years on, however, the complexity of the Internet, with its decentralized control and variability along every dimension, has meant that routing metrics are largely engineered by human operators on a relatively slow time scale. While there continue to be advances in online traffic engineering (e.g., [4]), one wonders how long it will be before Internet routing operates as a fully automated control system.

The subject of routing metrics brings us to our next paper, which focuses on a different setting than the Internet — multi-hop wireless networks:

- D. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A High Throughput Path Metric for Multi-Hop Wireless Routing", *ACM Mobicom 2003, San Diego, CA, USA, September 2003.*

Much prior work on routing in such networks has used hop count as the routing metric. However, the quality of a route is also affected by other factors, including the interference between hops, and the data rate and packet error rate on each hop. De Couto et al. provide an elegant framework to combine these into a single metric. They propose the expected transmission count (ETX) metric, which simply adds up the total number of packet transmissions required to deliver a packet end-to-end. Thus ETX combines the hop count and the packet error rate into a single metric, under the (pessimistic) assumption that all pairs of links interfere with each other. Subsequent work by Draves et al. [2] on the weighted cumulative expected transmission time (WCETT) metric extends this to incorporate the link data rate and also the use of multiple non-interfering channels. It is interesting to note that neither ETX nor WCETT is explicitly load-dependent.

Continuing with the wireless thread, we turn to our next paper, which is on the design of a medium access control (MAC) protocol for wireless networks:

- P. Karn, "MACA – A New Channel Access Method for Packet Radio", *9th ARRL Computer Networking Conference, London, ON, Canada, September 1990.*

Karn points out that carrier sense multiple access (CSMA), widely used in packet radio networks, focuses on interference at the sender whereas the real action in such networks is at the receiver. In other words, a transmission would be successful so long as the receiver is in the clear, even if the sender is not. This simple yet insightful observation led Karn to propose Medium Access with Collision Avoidance (MACA), which avoids carrier sensing altogether and instead uses a Request to Send/Clear To Send (RTS/CTS) mechanism to achieve collision avoidance. MACA has been influencial, with the 802.11 standard adopting the RTS/CTS mechanism. However, 802.11 also includes carrier sensing, since the sender needs to be in the clear to be able to receive link-level ACKs. Furthermore, while RTS/CTS is supported by 802.11 implementations, the overhead of this mechanism has meant that it is typically invoked only for large packets.

Let me now switch gears to a different topic — network measurement. This has been and continues to be an active area of research. However, the shelf life of work in this area tends to be short, given the rapid evolution of the Internet. Nevertheless, every so often there comes a paper that makes a lasting impression. I have picked two such papers here, the first of which is:

- J. Bolot, "End-to-End Packet Delay and Loss Behavior in the Internet", *ACM SIGCOMM 1993, San Francisco, CA, USA, September 1993.*

Bolot performed measurements to infer the packet delay and loss characteristics on a transcontinental link between the U.S. and Europe. He used simple yet effective methods, driven by a model of queuing behaviour, to tease apart meaningful information from noise. A great example of this is the use of a phase plot that charts successive RTT samples against each other rather than against time, as in a timeseries. The elegance of the methods is what, in my opinion, makes this paper a lasting contribution, long after the measurements themselves have lost meaning (indeed, the transcontinental link that was measured had a bandwidth of 128 kbps, meager by today's standards!).

The second measurement paper in the list is:

- C. Dovrolis, P. Ramanathan, and D. Moore, "What do packet dispersion techniques measure?", *IEEE Infocom 2001, Anchorage, AK, USA, April 2001.*

The packet pair technique is widely used to measure the bottleneck link capacity of a path. While it was always recognized that the measurements would be noisy due to cross traffic, the conventional wisdom was that such noise could be mitigated by employing a train (packet train) of large probe packets and picking the mode of the resulting measurements. Through careful modeling and analysis, Dovrolis et al. show that each of these assumptions is incorrect. Specifically, they show that the asymptotic dispersion rate of a long packet train, in the presence of cross traffic, is lower than the path capacity. These findings, which are contrary to intuition, highlight the importance of really understanding what is actually being measured.

Finally, we turn to three papers that are great examples of new insight into networking protocols and problems that had escaped the attention of previous researchers. The first of these is:

- S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson, "The End-to-End Effects of Internet Path Selection", *ACM SIGCOMM 1999, Cambridge, MA, USA, August 1999.*

While there had been prior work on characterizing Internet routing and separately on end-to-end network performance, the new insight in the work by Savage et al. comes from considering these together. Specifically, that the sub-optimality of Internet routing meant that one could do better by constructing an overlay path, i.e., taking a "detour". This simple idea has spawned off a large body of follow-up work on overlay routing, for performance optimization as well as other reasons such as improving robustness [1].

The next paper shows the familiar TCP in a new light:

- S. Savage, N. Cardwell, D. Wetherall, and T. Anderson, "TCP Congestion Control with a Misbehaving Receiver", *ACM Computer Communication Review, Vol. 29, No. 5, October 1999.*

TCP's congestion control mechanism is key to its operation and the health of the network as a whole. Indeed, malicious senders may manipulate their congestion control behaviour to gain an unfair advantage. However, as Savage et al. point out, even a TCP sender that is playing by the rules could be tricked by a misbehaving receiver into becoming aggressive. This is arguably more worrisome than a misbehaving sender since client hosts, which are often under the control of end users, function primarily as TCP receivers. This vulnerability to a misbehaving receiver arises from the loose nature of the contract between the TCP sender and the TCP receiver. A good example is optimistic ACKing, where a receiver could gain unfair advantage by ACKing data that it has not even received, without the sender being any wiser. Beyond these specific vulnerabilities of TCP, I believe that a key contribution of the paper is in opening our eyes to need for careful thought in the design of robust protocols.

Finally, we have the paper that introduced network coding:

- R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network Information Flow", *IEEE Transactions on Information Theory, Vol. 46, No. 4, July 2000.*

The extensive literature in network flow problems has used a fluid paradigm to model the flow of information. However, network coding shows that we can do better by combining (i.e., coding) information packets at the intermediate nodes. The impressive gains offered by network coding have led to an explosion of interest in this technique, with applications in unicast and multicast transport, content distribution, and wireless networks. However, in my view, the most astounding aspect of network coding is that while it takes just a few moments to demonstrate it in action through an example (such as the one shown in Figure 7 in the paper), network coding, in its general form, was discovered just a few years ago. This is a reminder to all of us that networking is a young field and you never know when you might hit upon a gem!

REFERENCES

[1] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient overlay networks. In *SOSP*, 2001.

[2] R. Draves, J. Padhye, and B. Zill. Routing in Multi-radio, Multi-hop Wireless Mesh Networks. In *MobiCom*, 2004.

[3] V. Jacobson. Congestion Avoidance and Control. In *SIGCOMM*, 1988.

[4] S. Kandula, D. Katabi, B. Davie, and A. Charny. Walking the Tightrope: Responsive Yet Stable Traffic Engineering. In *SIGCOMM*, 2005.

[5] K. K. Ramakrishnan, S. Floyd, and D. Black. The Addition of Explicit Congestion Notification (ECN) to IP , September 2001. RFC 3168.

[6] K. K. Ramakrishnan and R. Jain. A Binary Feedback Scheme for Congestion Avoidance in Computer Networks with a Connectionless Network Layer. In *SIGCOMM*, 1988.